

A Theory of (the Technological) Mind: Developing Understanding of Robot Minds

by

Kimberly A. Brink

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2018

Doctoral committee:

Professor Henry M. Wellman, Chair
Professor Margaret Evans
Professor Susan Gelman,
Professor Benjamin Kuipers

Kimberly A. Brink

kabrink@umich.edu

ORCID iD: 0000-0001-5947-2350

© Kimberly A. Brink 2018

DEDICATION

This dissertation is dedicated to the strong and wonderful women in my life who were there when I needed them. Thank you, Tyler, Emma, Margaret, Jenn, Shee Shee, Lisa, Steph, Katie, and April. Thank you, Ann and Heather. And thank you to my mom.

ACKNOWLEDGMENTS

This dissertation would not exist if not for Henry M. Wellman, my primary advisor, who entertained and nurtured my many interests. I will forever be indebted to him. I thank my committee members Margaret Evans, Susan Gelman, and Benjamin Kuipers for their invaluable feedback and investment in both me and my work. I also thank my collaborator, Kurt Gray, for his contributions to this work.

This research would not have been possible without the funding and training provided by the National Institute of Child Health and Human Development, the Rackham Graduate School, and the Center for Human Growth and Development.

Though this is my dissertation, the research reported here reflects the contributions of many research assistants with whom I have had the great pleasure to work. I thank those research assistants who dedicated their time and efforts to assist with these projects as well as related projects not reported on here. Countless thanks go to: Clare Anderson, Kaela Brown, Annie Dreisbach, Lorene Dreisbach, Makenzie Flynn, Kristen Goncalves, Sara Guttentag, Taryn Hayes, Alyssa Krentzel, Emily Kunster, Alysse Quinn, Caitlin Schneider, Rebecca Seguin-Skrabucha, Kayla Winer, Alexis Yezbick, and Lauren Yousif. A special thanks also goes to the laboratory coordinators who made this work possible: Cassie Neely, Hope Peskin-Shepherd, and Talie Toomajian.

In addition to the many hours invested by my research assistants, the research reported on here would not have been possible without the support of the University of Michigan Living Lab

Program. Thank you to Craig E. Smith for bringing the model to Ann Arbor. Thank you, also, to the University of Michigan Museum of Natural History (Kira Berman, Brittany Burgess) for generously supporting my work.

Outside the lab, I received excellent teaching mentorship from Drs. Ioulia Kovelman, Shelly Schreier, Rich Gonzalez, Daniel Weissman, and Matthew Diemer. From them, I learned not only how to teach, but how to mentor students and research assistants in ways that would promote their learning and nurture their curiosities. Thank you for helping me be a better teacher and researcher.

Throughout my time at Michigan, I also benefited from many great friendships. I thank my cohort for their helpful feedback and constant encouragement: Sammy Ahmed, Nkemka Anyiwo, Emma Beyers-Carlson, Fernanda Cross, Margaret Echelbarger, Arianna Gard, Amira Halawah, Tyler Hein, Jasmine Manalel, Lauren Tighe, and Neelima Wagley. I also thank my lab mates, past and present: Maria Arredondo, Lindsay Bowman, Amanda Brandone, Jasmine DeJesus, Rachel Fine, Selin Gulgoz, Danielle Labotka, Jonathan D. Lane, Kristan Marchak, Rebecca Marks, Ariana Orvell, Rachna Reddy and Steven Roberts. Beyond my cohort and the lab, I thank Paige Safyer, Jennifer Dibbern, Shen Huang, Shee Shee Jin, Lisa Johnson, Stephanie Oh, Jyotsna Venkataramanan, Katie Asher Halquist, Maggie Halpern, Nicholas Alderink, and April Rogers for their friendship.

To even get to this point, I am heavily indebted to several people who led me to this path. First, I thank Rebecca Saxe for igniting my passion for psychological research. Second, I thank Laura Schulz for introducing me to the incredible world of infant research. Third, I thank Luca Bonatti for taking a chance on a research assistant whom he'd never met to cross an ocean and work with babies. Fourth, I thank Heather Ann Thompson for inspiring and supporting my

pursuit of graduate school. Fifth, I thank Dillon Thompson Erb for guiding and supporting me toward this path. And sixth, I thank Henry M. Wellman, once more, for allowing me to pursue a novel and odd research topic with endless enthusiasm and steadfast support.

Lastly, I thank those closest to me who were there when times were light and dark. Thank you to Tyler, Jenn, and Katie, the incredible women who were on the other side of a phone call, cup of coffee, or glass of wine whenever I needed them. And thank you to my parents, Marty and Lisa Brink, and my big brothers, Matt and Nathan Brink, who all nurtured and encouraged my eccentricities. I could not have asked for a more supportive, loving, and wonderful family.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF APPENDICES	x
ABSTRACT	xi
CHAPTER	
I. Introduction	1
Children’s Developing Understanding of Robots	4
Research Questions	7
Outline of Research	8
II. Creepiness Creeps In: Uncanny Valley Feelings Are Acquired in	
Childhood	9
Abstract	9
Method	15
Results	21
Discussion	

III. Robot Teachers for Children? Young Children Trust Robots	
Depending on their Perceived Accuracy and Agency	36
Abstract	36
Study 1	39
Study 2	48
General Discussion	53
IV.A Theory of the Robot Mind: Developing Beliefs about the Minds	
of Social Robots	58
Abstract	58
Study 1: Adults	64
Study 2: Children	78
General Discussion	89
V. General Discussion	96
Related Research on Robot Instructors	98
Future Directions	100
REFERENCES	105

LIST OF TABLES

TABLE

II.1 Exploratory Factor Analysis of Interview Items	20
II.2 Regression Analyses Predicting Uncanniness Difference Scores	22
III.1 Social Robot Responses for Each Trial	42
III.2 Inanimate Machine Responses for Each Trial	51
IV.1 Proportion of Responses in Each Type of Mental Attribution	73

LIST OF FIGURES

FIGURE

I.1 An Array of Robots Designed Specifically to Interact with Children	4
II.1 A Schematic Depiction of the Theoretical Uncanny Valley	11
II.2 Still Frames from the Videos of Each Robot	13
II.3 Images to Assist Children in Answering Likert-Scale Type Questions	17
II.4 The Interaction Between Robot Type and Age	23
II.5 Measure of Uncanniness	25
II.6 Creepy-Weird Stimuli Used in Follow-Up Interview	27
III.1 Still Frame from Study 1	41
III.2 Familiar and Novel Objects Shown in Both Study 1 and Study 2	43
III.3 Proportion of Correct Responses to Accuracy, Ask, and Endorse Questions	46
III.4 Still Frame from Study 2	50
IV.1 Still Images of 10 Robots Presented in Study 1	62
IV.2 A Diagram of the CFA Model for Studies 1 and 2	69
IV.3 A Visualization of K-Means Clustering for Adult Responses	72
IV.4 The Additional Robot ASIMO, Included in Study 2	79
IV.5 A Visualization of K-Means Clustering for Child Responses	84

LIST OF APPENDICES

APPENDIX

II.A Robot Beliefs Interview	33
III.A Reduced Robot Beliefs Interview	56
IV.A Robot Beliefs Interview – Adults	93
IV.B Robot Beliefs Interview – Children	94

ABSTRACT

The purpose of this dissertation is to explore how children attribute minds to social robots and the impacts that these attributions have on children's interactions with robots, specifically their feelings toward and willingness to trust them. These are important areas of study as robots become increasingly present in children's lives.

The research was designed to address a variety of questions regarding children's willingness to attribute mental abilities to robots: (1) To what extent do children perceive that social robots share similarities with people and to what extent do they believe they have human-like minds? (2) Do attributions of human-like qualities to robots affect children's ability to understand and interact with them? (3) Does this understanding influence children's willingness to accept information from robots? And, of crucial importance, (4) how do answers to these questions vary with age?

Across a series of five studies, I investigated children's beliefs about the minds of robots, and for comparison adults' beliefs, using survey methods and video stimuli. Children watched videos of real-life robots and in response to targeted questions reported on their beliefs about the minds of those robots, their feelings about those robots, and their willingness to trust information received from those robots. Using a variety of statistical methods (e.g., factor analysis, regression modeling, clustering methods, and linear mixed-effects modeling), I uncovered how attributions of a human-like mind impact feelings toward robots, and trust in information received from

robots. Furthermore, I explored how the design of the robot and features of the child relate to attributions of mind to robots.

First and foremost, I found that children are willing to attribute human-like mental abilities to robots, but these attributions decline with age. Moreover, attributions of mind are linked to feelings toward robots: Young children prefer robots that appear to have human-like minds, but this reverses with age because older children and adults do not (Chapter II). Young children are also willing to trust a previously accurate robot informant and mistrust a previously inaccurate one, much like they would with accurate and inaccurate human informants, when they believe that the robot has mental abilities related to psychological agency (Chapter III). Finally, while qualities of the robot, like behavior and appearance, are linked to attributions of mind to the robot, individual differences across children and adults are likely the primary mechanisms that explain how and when children and adults attribute mental abilities to robots (Chapter IV). That is, individuals are likely to attribute similar mental abilities to a wide variety of robots that have differing appearances and engage in a variety of different actions.

These studies provide a variety of heretofore unknown findings linking the developmental attributions of minds to robots with judgments of robots' actions, feelings about robots, and learning from robots. It remains to be seen, however, the exact nature of the mechanisms and the child-specific features that increase children's willingness to attribute mental abilities to robots.

Chapter I

We live in a world increasingly filled with smart technology—laptops, tablets, smart phones, autonomous cars, and now robots. And, every day, these devices are behaving more and more like people, as if they have minds of their own. They can carry on conversations, answer questions, make jokes, and even make plans for how to get to the grocery store. New, socially interactive devices are emerging daily that push the boundaries of our understanding of what makes us human and what makes these devices machines.

The changing roles of technology pose a variety of intriguing questions concerning how our understanding of a human-like mind impacts our interactions with these devices: (1) To what extent do we perceive that these devices share similarities with people and to what extent do we believe they have human-like minds? (2) Do attributions of human-like qualities to these devices affect our ability to understand and interact with them? (3) Does this understanding influence our willingness to accept information from these devices? And, of crucial importance, (4) how do answers to these questions vary with age? Not only are these devices increasingly present in the lives of adults, but they are already interacting with children in their homes, schools, and hospitals—children are playing with and learning from these devices all over the world.

In this dissertation, I investigated these varieties of understanding and their impact on children's interactions with these devices, with a focus on children's feelings toward and trust in

social robots. Social robots are a special, and specially revealing, form of smart technology for both practical and substantive reasons.

Substantively, social robots provide a unique perspective with which to investigate children's conceptual development, specifically their developing concept of minds. Social robots are among a set of emerging socially interactive devices (e.g., computers, smart phones, autonomous cars) that appear to have minds of their own. They can simultaneously share characteristics with humans, like goal-directed behavior or a human-like appearance, while also sharing features with inanimate objects, like being built by humans and operating on electricity. Therefore, we can manipulate robots, as human-made machines, to dissect the mechanisms that allow children to attribute minds to these devices. By investigating children's understanding of the minds of robots, we can explore the nature of children's beliefs about the mind, the capacity to which children believe technology can have human-like minds, and the qualities that encourage children to attribute minds to technology.

Further, and of practical importance, children's understanding of robots is worthy of investigation because robots are becoming increasingly present in the lives of children. Robots are now being designed to befriend, teach, and care for our children. Dozens of robots have been released over the past few years equipped specifically to interact and play directly with children. iPal, Jibo, and Zenbo (see Figure I.1), all Pixar-like robots, are designed to play games, answer questions, read stories, and even watch children unsupervised (Glaser, 2016; Low, 2016; Wong, 2016). Moreover, several robots have been working with children in classrooms, daycares, clinics, and hospitals for years. Across the globe, robots are teaching children language skills (Movellan, Eckhardt, Virnes, & Rodriguez, 2009), physical exercises (Mejías et al., 2013), and social skills (Ricks & Colton, 2010). Investigations into children's understanding of robots

becomes more important as children are increasingly expected to interact directly with these devices. And of particular focus in this dissertation, I explore how perceptions of mind impact children's feelings toward and willingness to trust robots.

Children's feelings toward robots. One important outcome of perceiving a human-like mind in a technological device is its impact on comfort with that device. For adults, the perception of a human-like mind in a robot has been linked to feelings of discomfort with that robot (K. Gray & Wegner, 2012). This has been found for adults but not yet for children. By investigating how children's attributions of a human-like mind impact their feelings toward robots, we can more fully understand the importance of a human-like mind for child-robot interactions; for example, we can more fully understand the origins of discomfort with human-like technologies. As robots become increasingly present in the lives of children, it is crucial that we understand the factors that increase or decrease children's comfort with these devices. Engineers and designers hope that children will play and learn from robots. If a child finds a robot unsettling, this will likely hinder any intended positive outcomes from interactions with that robot.

Children's willingness to trust robots. Moreover, in their roles as educational devices, social robots provide a unique means with which to investigate how a human-like mind may impact children's learning more generally. Until recently, children have efficiently and effectively learned about their world through knowledge passed on by other people (e.g., parents, teachers, peers). This phenomenon, recently studied mostly under the headings of "trust in testimony" or "natural pedagogy," shows how children are adapted to learn general knowledge from human communication (Csibra & Gergely, 2009). As robots become increasingly human-like, they may be able to take advantage of this form of social learning to effectively transmit

information to children. More and more, we are asking children to accept these technologies and embrace the information that they receive from them. And yet, there has been limited focus on how and whether children will accept and learn from these devices and the sorts of features that can enhance or detract from such interactions.

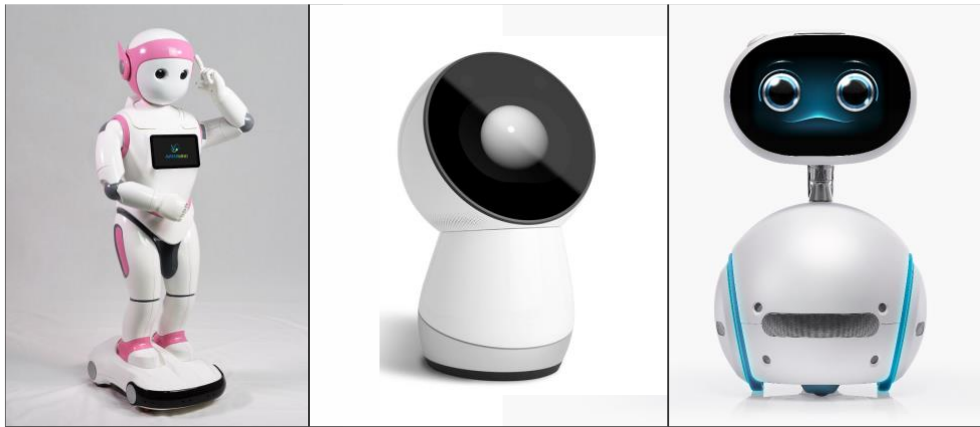


Figure 1.1. An Array of Robots Designed Specifically to Interact with Children. An array of robots designed specifically to interact with children: iPal (left panel), Jibo (center panel), Zenbo (right panel).

The importance of child-focused research. Lastly, until recently, most research on the effectiveness of robotic social companions has focused primarily on features of the robot. I argue, however, that research is badly needed to assess how children’s cognitive abilities, their developmental trajectories, and the design of the robot work together to impact children’s learning and feelings towards robots. Thus, I propose possible answers as to *how* and *why* these child factors—cognitive abilities that themselves change with age—interact with the overt features of social robots to affect the quality of social robots and children’s trust of and learning from them. Specifically, I demonstrate how children’s understanding of the mental abilities of robots directly impacts their relationships with them.

Children’s Developing Understanding of Robots

It is relatively easy to imagine ways that children's own understanding of robots could impact their feelings about them and relatedly their willingness to trust and learn from them. Children may recognize that there are similarities between robots and other familiar categories (i.e., people, animals, or artifacts) and feel comfortable with them. Contrastingly, children may perceive robots as altogether unfamiliar and unique categories that either require further exploration or elicit distrust.

Furthermore, these expectations about robots could, in turn, affect children's willingness to learn from and trust robots. If children perceive robots as similar to infallible tools, like calculators or dictionaries, they may indiscriminately accept information from all robots. Or if children perceive robots, especially humanoid robots, as similar to people with thoughts, emotions, and sometimes flawed beliefs about the world, they may differentially accept information from only trustworthy ones. Alternatively, if a child perceives robots as entirely different from humans and other common sources of information, they may ultimately consider robots to be suspect and never trust information from them. To embrace and learn from robots, children likely must first determine what they believe about the minds of robots: are they infallible tools, intelligent beings, or something else entirely?

The nature of robots, however, is not easy to define. Robots are unique devices that can simultaneously share similarities with artifacts, animals, and even humans. They, like other artifacts, are designed and built by humans. They do not live, grow, breathe, or (as adults typically believe) experience feelings (K. Gray & Wegner, 2012). Therefore, children may perceive robots as artifacts or infallible tools similar to textbooks or other educational devices. Yet, unlike more common artifacts such as books, bikes, beds, and balls, social robots are not only human-built but human-like to varying degrees. They can look, behave, and at times even

“think” like humans or animals. Hence, children may alternatively see them as intelligent beings that can think and reason. As such, robots, like humans, might be fallible and should be considered carefully before trusting them. Indeed, research on children’s developing understanding of robots reflects children’s attempts to rationally and emotionally embrace/resist these competing identities of robots.

Research on early childhood understanding of robots suggests that children initially attribute a myriad of human-like mental abilities to robots. For instance, they expect robots to have emotional, social, as well as perceptual abilities. Young children report that robots have perceptual abilities, like sight and touch: 3-year-olds claimed that a robot dog could see and be tickled (Jipson & Gelman, 2007). Nine- and 12-year-olds similarly reported that a 3-foot tall interactive robot, Robovie, could be intelligent, have interests, and experience emotions (Kahn et al., 2012). These children also believed that Robovie could be their friend and could comfort them if they were sad. Whereas young children appear to treat and think about robots like people or animals, they, however, do not equate robots with them. Children recognize that a robot dog is not identical to a real dog as demonstrated by their claims that robot dogs do not have biological qualities (Melson, Kahn, Beck, & Friedman, 2009): young children report that robotic dogs cannot grow or eat like real dogs for example (Jipson & Gelman, 2007).

As children age, however, their beliefs about robots change—expectations that robots have emotional, social, and perceptual capacities decrease. Older children are less likely to report that a robot has emotions, desires, or is capable of autonomous action (Mikropoulos, Misailidi, & Bonoti, 2003). Five-year-olds were less likely to claim that a robot dog could think or feel happy compared to 3-year-olds (Jipson & Gelman, 2007). Children older than 7 spoke differently about robots, more often using language specific to man-made machines, compared to children

younger than 7 (Okita, Ng-Thow-Hing, & Sarvadevabhatla, 2011). Fifteen-year-olds were also less likely to believe that Robovie could have interests, experience emotions, or be a friend compared to 9- and 12-year-olds (Kahn et al., 2012). We similarly show this relation in our own research (described in this dissertation): children's reports of robots' perceptual abilities and psychological abilities decline across 3 to 18 years of age. With age, children begin to dissociate psychological, emotional, social, and perceptual abilities from robots and recognize that robots are more similar to artifacts.

Finally, by adulthood, beliefs about the capacities of robots have decreased to a substantially reduced set of expectations. Adults essentially expect that robots are only capable of some forms of thinking and decision-making, and they deny robots the ability to feel pain or fear (K. Gray & Wegner, 2012).

Given this sort of evidence, I argue that there is a transition between how children and adults think about robots. Young children attribute more social, psychological, and perceptual abilities to robots. Yet, with age and experience, children gradually adjust their beliefs about robots so that, by adulthood, robots possess only limited psychological agency, the ability to think and make decisions. I argue that this change in understanding about the minds of robots impacts children's interactions with and feelings toward robots. My research, described next, speaks to this transition and its impact on children's feelings toward robots.

Research Questions

This focus on social robots allows me to carefully and precisely address the questions presented above: (1) To what extent do children perceive that social robots share similarities with people and to what extent do they believe that they have human-like minds? (2) Does an understanding that robots can and do have human-like minds affect children's feelings toward

these robots? (3) Does this understanding also influence children's willingness to accept information from robots? And, finally, (4) how do answers to these questions vary with age?

Outline of Research

The studies I present here are a concerted effort to address these questions, specifically exploring how children's attributions of mind to social robots impact their affinities for robots and their willingness to learn from them.

The first study (Chapter II) investigates the extent to which children perceive that robots have human-like minds and how those perceptions impact their feelings toward robots. I also investigate how these attributions and their relation to feelings toward robots changes with age. This research was published in *Child Development* (Brink, Gray, & Wellman, 2017).

The second set of studies (Chapter III) encompasses two studies that investigate the extent to which young children accept information from robots. Moreover, it examines how children's attributions of mind to robots influence their willingness to trust robots (Brink & Wellman, submitted). These studies have been submitted for publication at *Child Development*.

The third set of studies (Chapter IV) more broadly explores how adults and children attribute a human-like mind to a vast array of robot features. The first of two studies here is complete with data collected from over 400 adult participants. The second study is still in progress, I provide a preliminary report for it based on data from over 100 child participants.

Chapter II

Creepiness Creeps In: Uncanny Valley Feelings Are Acquired in Childhood

Abstract

The Uncanny Valley posits that very human-like robots are unsettling, a phenomenon amply demonstrated in adults but unexplored in children. 240 3- to 18-year-olds viewed one of two robots (machine-like or very human-like) and rated their feelings toward (e.g., “Does the robot make you feel weird or happy?”) and perceptions of the robot’s capacities (e.g., “Does the robot think for itself?”). Like adults, children older than 9 judged the human-like robot as creepier than the machine-like robot—but younger children did not. Children’s perceptions of robots’ mental capacities predicted uncanny feelings: children judge robots to be creepy depending on whether they have human-like minds. The uncanny valley is therefore acquired over development and relates to changing conceptions about robot minds.

Keywords: uncanny valley, theory of mind, social cognition

Creepiness Creeps In: Uncanny Valley Feelings Are Acquired in Childhood

All day, every day, both children and adults try to get inside the minds of others, wondering about their thoughts, feelings, and intentions. Until the past few decades, these minds have been those of flesh and blood—humans and animals—but now we are faced with minds made of metal and silicon, including smart phones and cloud computing. How do we learn to make sense of these artificial minds?

Nowhere is this question more pressing than with robots, who have self-directed mechanical minds dwelling inside human-like bodies. The National Robotics Initiative foresees a future in which “robots are as commonplace as today's automobiles, computers, and cell phones. Robots will be found in homes, offices, hospitals, factories, farms, and mines; and in the air, on land, under water, and in space (National Robotics Initiative 2.0, 2017). In fact, robots are already entering homes, not only to help adults with household chores, but also to play with and teach children. Moreover, several robots have been working with children in classrooms, daycares, clinics, and hospitals for years. Robots are teaching children language skills (Movellan et al., 2009), mathematics (Wei, Hung, Lee, & Chen, 2011), science (Hashimoto, Kobayashi, Polishuk, & Verner, 2013), physical exercises (Mejías et al., 2013), and even social skills (Ricks & Colton, 2010). Dozens of robots have been released in the past year alone designed specifically to interact with children. As robots become increasingly present in our lives and the lives of our children, it becomes more and more important to explore how we reason about the minds of these devices and how this reasoning impacts our interactions and feelings toward them.

Work with adults has identified one phenomenon, in particular, that could shed light on this topic. Specifically, decades of research reveal that while adults prefer robots that are

somewhat human-like, they find very human-like robots unnerving—the “uncanny valley” phenomenon (MacDorman, Green, Ho, & Koch, 2009; Mori, MacDorman, & Kageki, 2012). According to theories of the uncanny valley, machines become increasingly attractive as they become more human-like until they reach a threshold at which they become too human-like and are considered uncanny and creepy (see Figure II.1). This *dip* in affinity for very human-like robots is the uncanny valley. Closely human-like robots are distinctly creepier than other robots and, in particular, creepier than the more unsettling of machine-like robots. Support for the uncanny valley comes from many studies in which adults report feeling greater unease when presented with extremely human-like robots compared to others (K. Gray & Wegner, 2012; MacDorman, 2006).

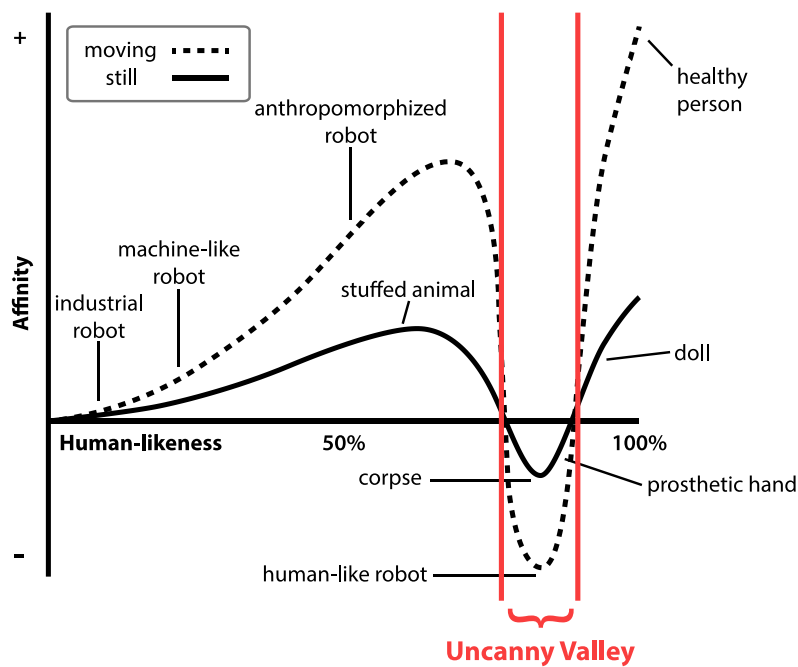


Figure II.1. A Schematic Depiction of the Theoretical Uncanny Valley. A schematic depiction of the theoretical Uncanny Valley (figure closely derived from Figure 2 in Mori et al., 2012). The uncanny valley is defined as the precipitous dip in affinity for closely human-like robots.

Two theories have been proposed to explain the uncanny valley's origins. One references innate evolutionary concerns (Steckenfinger & Ghazanfar, 2009) including the innate drive to avoid illness. Human-like robots may display "visual defects" that are interpreted as signs of a "communicable disease" thus producing a creepy response. Alternatively, innate face processing mechanisms may recognize visual defects in very human-like faces compared to real human faces and thus process those faces as unattractive and creepy. Various facial processing mechanisms and standards are in fact apparent even in infants (see Langlois, Roggman, & Rieser-Danner, 1990, for evidence that infants prefer attractive faces). This evolutionary account receives further support from research demonstrating that even monkeys experience an uncanny valley when viewing computer simulated images of monkey faces (Steckenfinger & Ghazanfar, 2009).

An alternative theory proposes that for humans the uncanny valley is not simply a by-product of evolutionary perceptual responses but instead depends on an acquired everyday understanding of what makes humans distinct from machines (MacDorman & Ishiguro, 2006). Feelings of uncanniness may instead emerge when a human-like machine violates our learned expectations of how a machine should look or behave. In the case of robots, for example, when a machine closely resembles a thinking and feeling human being, this would violate our expectations that machines should be inanimate and hence incapable of thought and experience. Specifically, a very human-like appearance in a machine can prompt attributions of a human-like mind (Epley, Waytz, & Cacioppo, 2007), and as human-like minds are seldom ascribed to robots (H. M. Gray, Gray, & Wegner, 2007), this mismatch causes feelings of uncanniness (K. Gray & Wegner, 2012). Indeed, research with adults reveals that the more robots are seen to have human feelings, the more unnerving they seem (K. Gray & Wegner, 2012). Violations of expectations

about the behavior and appearance of machines and humans thus link to the uncanny valley phenomenon in adults.

While the uncanny valley has been studied in adults, its origins have never been studied in children. If evolutionary in nature, the uncanny valley should be evident in even the youngest children. However, if it is related to developing expectations about humans and machines, then it should emerge throughout childhood—perhaps in tandem with exclusive attributions of human-like minds to humans. As the origins and mechanisms of the uncanny valley have yet to be tested in children, we examine them here.

We offer a detailed look at the uncanny valley across development by measuring uncanny responses to videos of robots in children from ages 3 to 18. We used stimuli previously validated with adults (K. Gray & Wegner, 2012): videos of the same robot that revealed either its machine-like or human-like nature (see Figure II.2). We showed these videos to children and then assessed their feelings of creepiness and also their attributions of mind—thinking (agency) and feeling (experience)—toward the robots. By assessing feelings of unease and mind attribution across a large age range, we could detect whether (and when) the uncanny valley develops and its potential link to children’s understandings of robot minds.



Figure II.2. Still Frames from the Videos of Each Robot. Still frames from the videos of each robot: Kaspar from the back (left panel), Kaspar from the front (center panel) and Nao (right panel).

We expected one of three possible patterns would likely appear: (1) the uncanny valley would be present at even the youngest ages, offering support for the evolutionary perspective. (2) The uncanny valley would emerge in early childhood in tandem with general perceptions of mind—offering support for the developmental perspective. (3) The uncanny valley would emerge in later childhood, when children develop more sophisticated—and specific—understandings of the different kinds of minds possessed by machines, also supporting a developmental perspective.

The evolutionary perspective suggests that even the youngest children should find human-like robots more unnerving than machine-like robots, *irrespective* of attributions of mind. On the other hand, the developmental perspective suggests that the uncanny valley will emerge in childhood, perhaps even early childhood, when children have begun distinguishing humans (and human minds) from other categories. Theory-of-mind research shows that 3- to 5-year-olds become quickly adept at attributing mental states such as beliefs and desires to humans (see meta-analyses by Milligan, Astington, & Dack, 2007; Wellman, Cross, & Watson, 2001). This too might predict that the uncanny valley would be evident in our youngest age group. However, this possibility could, nonetheless, be distinguished from the evolutionary one if the presence of uncanny valley responses in early childhood is related to children’s attributions of mind.

Alternatively, the uncanny valley could instead arise around middle childhood—when children develop richer understandings of folk biology and folk psychology, and begin to separate the concepts of minds, brains, bodies and machines (Wellman, 2014). For example, it is only at about 9-12 years that children truly understand differences between the mind (as more “mental”) and the brain (as more part of the biological body; C. N. Johnson & Wellman, 1982; Richert & Harris, 2006). This understanding that the mind stems from the biological brain (i.e., a

neuro-physiological “machine”) could support the development of the uncanny valley: the uncanny valley may result from the mismatch of perceiving a human mind as stemming from a machine brain. Indeed, post-preschool children, as they age, expect machines to have fewer mental abilities (Kahn et al., 2012 which examined children age 9 to 15). It may not be until this later age that children develop an understanding that robots, as machines, should not have minds as humans do, making them uncanny when they seem like they do.

Method

Participants. 240 children (117 females), 3 to 18 years old, were recruited from a local natural history museum (218 children) or from a participant database (12). Children were questioned in a semi-isolated, quiet space within the museum or (for 12) in an on-campus laboratory space. One child was excluded due to incorrect parental report of their birthdate. Our sample was twice the number of participants ($N = 120$) used in a similar previous task (K. Gray & Wegner, 2012; Study 1). Power analyses indicate that $N = 240$ exceeds .80 statistical power (Cohen, 1988).

Because data were collected in a public space, we did not collect information regarding children’s race, ethnicity, or socioeconomic status. Written parental consent and verbal child assent were obtained first; children received a small toy for participating.

Videos. Children were randomly assigned to watch short videos of either a closely human-like robot or a more machine-like robot (Figure II.2), the two used in K. Gray & Wegner (2012). For the human-like robot, 119 children watched 16s of Kaspar moving its head, filmed from the front with its human-like face clearly visible. In the machine-like robot condition, 120 children watched 16s of the robot Kaspar moving its head while filmed from behind, where *only* its wiring and electrical components could be seen, no human-like features were visible.

Respondents could not infer that these views were of the same robot because (paralleling K. Gray & Wegner, 2012) these videos were presented between-subjects. This focal comparison controls for many irrelevant differences between the human-like and machine-like robots because the videos contain the same robot making the same movement but from different views (front vs. back).

Whereas our focal contrast compares the human-like and machine-like versions of Kaspar, we also included a video of Nao—a commercially available abstractly humanoid robot—as a baseline condition (See Figure II.2). We filmed Nao to mimic the human-like robot—only head, face and torso visible, moving its head from side to side, with no changes in expression. After viewing either the machine-like robot or human-like robot, 234 children watched the 16s video of Nao. We implemented this baseline condition using Nao as one indicator that children used our rating scale and terms appropriately (see below).

Nao has been used in previous studies with children (4 to 9 years) to effectively comfort them during stressful events (e.g., receiving a vaccination; Beran, Ramirez-Serrano, Vanderkooi, & Kuhn, 2013) and so is presumably not creepy or uncanny. Nao is also unlikely to be considered creepy because it resembles the friendly, animated robot protagonists portrayed in children’s films like WALL·E (Stanton, 2008) and Baymax in *Big Hero 6* (Hall & Williams, 2014). Thus, if children appropriately use our scale, they should provide low ratings of uncanniness for Nao. Whereas we expect that Nao should be rated low on feelings of uncanniness, we did not use Nao as an indicator of the presence or absence of the uncanny valley. First, Nao has never been empirically placed on the hypothesized Uncanny Valley gradient. Moreover, a contrast between Nao and the very human-like robot would be insufficient evidence to prove the existence of the uncanny valley. The uncanny valley is more specifically

defined as the dip in affinity when very human-like robots are perceived as creepier than even somewhat creepy machine-like robots. Finding that a very human-like robot (Kaspar from the front) is creepier than a not very creepy, and even comforting, humanoid robot (Nao) would not sufficiently demonstrate the presence or absence of the uncanny valley.

Task and design. After viewing each robot (machine- or very human-like, then Nao) children answered multiple questions presented in a two-part format. In the first part, children chose one of two options. For example, when asked “Does the robot think for itself?”, children answered either yes or no verbally or by pointing to a “thumbs up” (yes) or a “thumbs down” (no) card (see Figure II.3). If children answered yes, they then answered a second Likert-type scale question. For example, “How much does the robot think for itself?”: “a little bit,” “a medium amount,” or “a lot.” Children could answer verbally or by indicating on a scale with increasingly tall bars (see Figure II.3).



Figure II.3. Images to Assist Children in Answering Likert-Scale Type Questions. Images that were shown to children to aid them in answering the two-part survey questions: thumbs-up (yes), thumbs-down (no), and a scale with bars increasing in height (“a little bit,” “a medium amount,” or “a lot”). These exact depictions were taken from (Severson & Lemm, 2016).

Robot Beliefs Interview. Children were assessed on their feelings of uncanniness via two questions gauging the extent to which children felt the robot was creepy or unsettling: (1) “Do you feel the robot is nice or creepy?” If children reported the robot was creepy, we asked, “How creepy do you feel it is?” (2) “Does the robot make you feel weird or happy?” and, “How weird does it make you feel?” This question format resulted in a 4-point scale for each question coded as Nice/Happy (0), Creepy/Weird-a little bit (1), Creepy/Weird-a medium amount (2), and Creepy/Weird-a lot (3).

Then children answered eleven additional questions, ten of them addressing the robots’ mental capacities (adapted from H. M. Gray et al., 2007; Severson & Lemm, 2016)(see Appendix II.A for complete interview). Previous interviews with adults have identified components of mental capacity labeled “agency” and “experience.” Questions were designed to encompass similar factors in our sample: psychological agency (does the robot “do things on purpose?”, “choose to move?”, “think for itself?”, “know the difference between good and bad?”) and perceptual experience (would the robot “feel pain?”, “feel scared?”, “feel hungry?”). The same two-part question format resulted in a 4-point scale for each of these “mind” questions coded as No (0), Yes-a little bit (1), Yes-a medium amount (2), and Yes-a lot (3).

Procedure. Children were instructed that they would view videos of robots and answer questions about them. Children then answered three warm-up questions and were randomly assigned to watch a video of either the closely human-like or machine-like robot on an iPad. After watching the video, children completed the *Robot Beliefs Interview*. The video was paused so that a still frame of the robot was visible during the interview. Upon completion of the interview, children watched a video of Nao and performed the *Robot Beliefs Interview* once more for Nao while a still frame of Nao remained visible on the iPad.

Data analysis. *Exploratory factor analysis.* It was unclear whether children's perceptions and feelings toward robots would reveal the same factor structure as past work with adults, so we performed an exploratory factor analysis (EFA) separately for each of the two conditions (very Human-like and Machine-like). Using an oblique rotation, Kaiser's criterion (eigenvalues ≥ 1), a scree plot, and model fit indices (Preacher & MacCallum, 2003), three factors were identified: Uncanniness (machine-like: $\alpha = .62$, human-like: $\alpha = .75$), Agency ($\alpha = .72, .64$), and Experience ($\alpha = .73, .85$). Five additional items were pruned (see Appendix II.A) that had low factor loadings, cross-loaded on to multiple factors, or were not easily interpretable; these items were not included in the final factor analysis.

As shown in Table 2.1, the three factors (covering 8 items)—Uncanniness, Agency, and Experience—had identical factor structures across the two conditions (very human-like and machine-like) and provided high overall fit. Table 2.1 reveals that all items had a loading of at least .40 on their respective factors. Model fit indices also support the three-factor solution within each condition. For the machine-like condition, chi-square goodness of fit, $\chi^2(7) = 3.64$, $p = .82$, RMSEA = .00, 90% CI = [.00, .07], and TLI = 1.07, were all within their established cutoff ranges (Hu & Bentler, 1999). For the human-like condition, model fit indices approximated or were within established cutoff ranges: chi-square goodness of fit, $\chi^2(7) = 10.53$, $p = .16$, RMSEA = .07, 90% CI = [.00, .14], and TLI = .953. Cronbach's α ranged from .62 to .85 across all three factors and both conditions.

Table II.1

Exploratory Factor Analysis of Interview Items			
Items	Factors		
	Uncanniness	Agency	Experience
Do you feel the robot is creepy?	.81, .99	-.05, .02	.01, -.07
Does the robot make you feel weird?	.57, .55	.13, -.18	-.01, .12
When the robot moves, does it choose to move?	.08, .10	.59, .69	.06, .07
Does the robot think for itself?	.01, -.04	.77, .53	-.04, .34
Does the robot know the difference between good and bad?	-.14, -.10	.66, .47	.02, -.08
Would the robot feel pain?	-.03, .02	-.02, -.04	.93, .86
Would the robot feel scared?	.01, .03	.27, .03	.40, .82
Would the robot feel hungry?	.03, -.09	.27, .06	.51, .76
α	.62, .75	.72, .64	.73, .85

Note: Using an oblique rotation, Kaiser’s criterion (eigenvalues ≥ 1), a scree plot, and model fit indices, three factors were identified. The first nine rows represent the factor loadings for each item and, in the bottom row, the α values for each factor. Bolded numbers identify the items that were used to calculate the aggregates for their respective factors. The first number in each cell represents values for the machine-like robot and the second for the human-like robot respectively (as indicated by the two numbers in each column).

Attributions of Mind score. *Agency* was measured by averaging the items “does the robot choose to move?”, “think for itself?”, “know the difference between good and bad?” Thought, decision making and morality have been linked to psychological agency for adults (H. M. Gray et al., 2007). *Experience* was measured by averaging the items: “would the robot feel pain?”, “feel scared?”, “feel hungry?” These items also resembled those for perceptual experience in adult research (H. M. Gray et al., 2007). The aggregates for Agency and Experience were highly correlated, $r(236) = .49$. Thus, for conceptual reasons and to avoid issues of multicollinearity, Agency and Experience were averaged to create a composite measure of Attributions of Mind. This approach has also been used in the adult literature (see e.g., K. Gray, Knickman, & Wegner, 2011).

Uncanniness difference scores. To account for individual differences in children's use of the scale, the dependent variable of Uncanniness was converted to a difference score. This difference score was calculated by subtracting the uncanniness score for the baseline condition, Nao, from the uncanniness score for the focal robot that a child viewed. For example, for children that viewed the human-like robot, their uncanniness score for Nao was subtracted from their uncanniness score for the human-like robot. This difference score allows the comparison between our focal two robots to be individualized to the extent that we use Nao as a baseline, anchoring the score of the focal robots on an empirically verified comforting robot, Nao (Beran et al., 2013). This score thus also provides a control for children's unfamiliarity with robots in general which can vary from child to child and age to age. (See, Dalecki & Willits, 1991, for an explanation and justification of the statistical advantages of such comparisons.) We did not use Nao as a regressor in our subsequent regression analyses because preliminary analyses showed that feelings of uncanniness for Nao did not significantly predict feelings of uncanniness for the focal robots. Uncanniness difference scores will be referred to as Uncanniness scores.

Results

The Development of the Uncanny Valley

Attributions of Mind, Robot Type (very human-like vs. machine-like), and Age as well as interactions between Mind and Age, and Robot and Age were entered into a regression analysis predicting Uncanniness scores. As shown in Table 2.2, there were significant associations between Uncanniness scores and attributions of Mind and Robot Type, qualified by the interaction between attributions of Mind and Age and the interaction between Robot Type and Age. There was no main effect of Age.

Table II.2
Regression Analyses Predicting Uncanniness Difference Scores

	Estimate	Std. Error	β	t	p
Human-Like Robot	.32	.15	.14	2.10	.037*
Mind	-.25	.11	-.21	-2.31	.022*
Age	-.23	.14	-.20	-1.68	.094
Human-Like Robot x Age	.32	.15	.13	2.08	.039*
Mind x Age	.20	.10	.17	2.04	.043*
Intercept	.89	.12		7.41	<.0001***

Note: Robot type, attributions of Mind, the child’s age, and two interaction terms were entered into a regression analysis to predict Uncanniness difference scores. Attributions of Mind and age were centered. Robot type, attributions of Mind, the interaction between robot type and age, and the interaction between attributions of Mind and age all predicted reports of uncanniness, $R^2 = .12$. * $p < .05$. *** $p < .0001$.

The interaction between Robot Type and Age indicates that the uncanny valley develops. The positive interaction indicates that as children age, the human-like robot is perceived as increasingly creepier than the machine-like robot. A plot of the interaction can be seen in Figure II.4. Tests of simple slopes indicated that the human-like robot does not become creepier than the machine-like robot until approximately 9 years of age. Before 9 years of age, Robot Type did not predict feelings of uncanniness (i.e., both the human-like and machine-like robot were equally creepy): at age 4, $\beta = -.08$, $p = .57$ and at age 8, $\beta = .13$, $p = .11$. At 9 years, however, the uncanny valley effect emerges with Robot Type significantly predicting uncanniness, $\beta = .18$, $p = .02$, where the human-like robot is significantly more uncanny than the machine-like robot. The uncanniness of human-like robots relative to machine-like robots continues to increase up to 16 years, $\beta = .55$, $p = .002$.

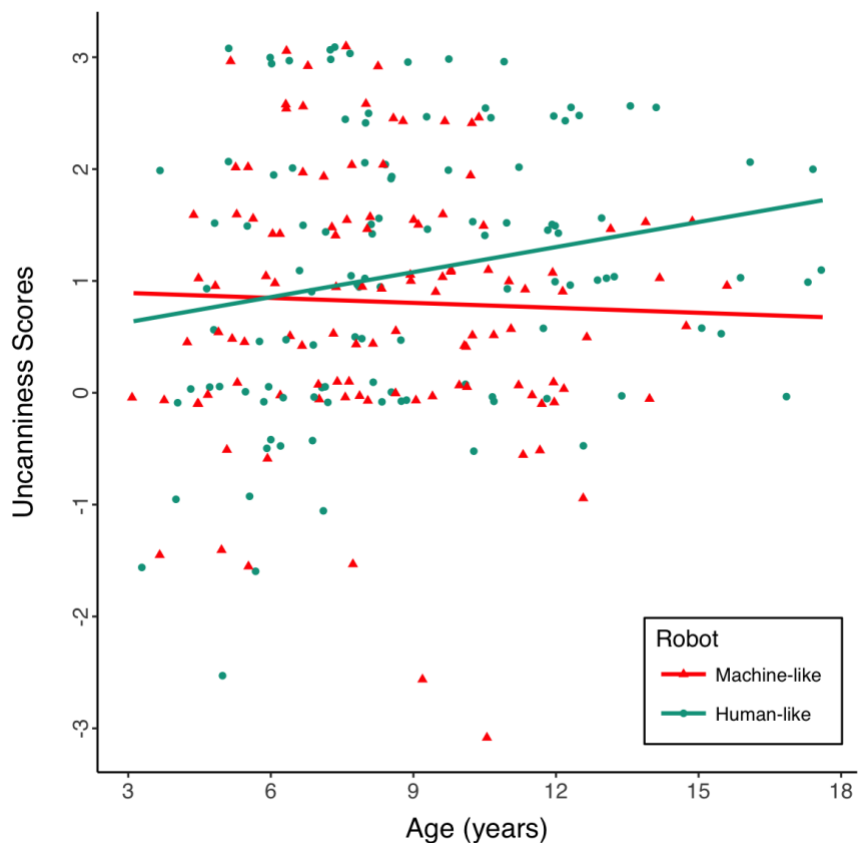


Figure II.4. The Interaction Between Robot Type and Age. The interaction between Robot Type and Age, $\beta = .14$, $t(223) = 2.08$, $p = .04$, shows that the uncanny valley develops. The positive interaction indicates that as children get older the human-like robot becomes increasingly creepier than the machine-like robot as shown here. The development of the uncanny valley effect is demonstrated by the increasing distance between the two lines (machine-like vs. human-like) with age.

Uncanniness and Mind

The development of the uncanny valley is also linked to attributions of Mind, shown by the significant interaction between attributions of Mind and Age on Uncanniness scores (as shown in Table 2.2. This interaction indicates that, as children get older, the association between attributions of mind and feelings of uncanniness changes. Specifically, tests of simple slopes show that, in young children, increased attributions of mind tend to predict *decreased* feelings of uncanniness in young children: for children ages 4 to 9 attributions of mind negatively correlated

with the uncanny response—at 4 years, $\beta = -.55, p < .001$, and at 9 years, $\beta = -.23, p = .04$. For older children, ages 10 to 18, this negative correlation started to disappear and began to trend positive, although not significantly so—at 10 years, $\beta = -.16, p = .21$, and at 16 years, $\beta = .23, p = .42$. The upper age limit of our sample however had fewer participants and therefore less power to test for a statistically positive association. This descriptively positive trend at the oldest range of our child sample fits qualitatively with findings in the adult literature in which feelings of uncanniness become positively associated with attributions of mind (K. Gray & Wegner, 2012). In total, this significant interaction suggests that the emergence of the uncanny valley is associated with children's perceptions of mind, particularly for younger children.

Uncanniness Responses for Nao

Figure II.5 shows the raw scores for feelings of uncanniness for all three robots. Inspection of this figure shows that Nao was consistently rated low on uncanniness across all ages in our sample. A regression analysis predicting raw uncanniness scores while controlling for age and comparing Nao with the least creepy of the two focal robots, the machine-like robot, showed that Nao is less creepy than the machine-like robot, $\beta = -.82, t(110) = -9.34, p < .0001$. Nao is also less creepy than the closely human-like robot, $\beta = -1.05, t(111) = -11.58, p < .0001$. Statistically, appropriate baseline conditions should have low, stable scores on the variable of interest, thus dispelling the possibility of a yes bias. Our expectation for Nao to be minimally uncanny (as explained above) was thus confirmed empirically, a result that supports its use as a baseline condition for creating Uncanniness difference scores. Further, raw uncanniness scores for Nao did not differ between children who first saw the human-like ($M = 1.31, SD = .67$) or first saw the machine-like robot ($M = 1.36, SD = .69$), $t(224.25) = -.53, p = .59, 95\% CI = [-.23,$

.13]. Because the robot that children saw first did not impact their responses to Nao, this provides additional empirical justification for its use as a baseline condition.

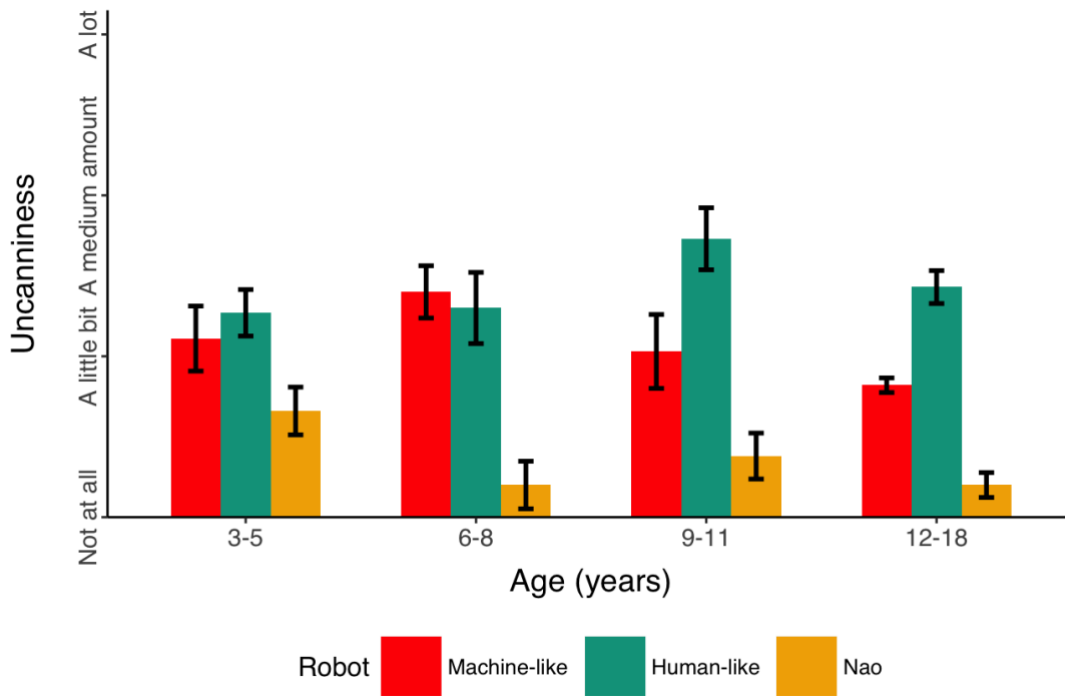


Figure II.5. Measure of Uncanniness. Measure of uncanniness, the aggregate of the two questions that measured whether children perceived the robot to be creepy or weird, for each of the three robots: the machine-like robot (Kaspar from the back), the human-like robot (Kaspar from the front), and Nao. Error bars represent standard errors.

Appropriate Understanding of “Creepy” and “Weird”

Methodologically, the comparison between Nao and the other robots shows that even the youngest children respond to our uncanny valley questions and test format with varied answers across the conditions, indicating that they offered meaningfully differential responses; children did not merely demonstrate a yes bias to our scale. Although their uncanny ratings, based on use of the terms “creepy” and “weird,” do not distinguish the machine-like and the human-like robot, they do distinguish between Nao and these two.

For further confirmation of their understanding of these two key terms, we tested a separate sample of 20 young children on their understanding of the terms “weird” and “creepy” via a brief interview. These children came from the same local population as those in our main study and were equivalent in age to children at the lower end of the age range in our sample ($M = 3.40$ years, $SD = .34$, range = [3.01, 3.99]). These children were presented with two paired images of a typical toy (i.e., a stuffed giraffe or tricycle) and a clearly strange toy (toys modeled after the creepy/weird toys in Toy Story; Lasseter, 1995) shown in Figure II.6. For one pair, children were asked to select which toy was creepy and, for the other pair, which toy made them feel weird (see questions in Figure II.6). On these tasks, 95% of these young children appropriately chose the strange toy as creepy and 85% chose the other strange toy as making them feel weird. Overall children were 90% correct on these items that used terms and phrasing closely similar to those in the two items that constituted our uncanny index. Thus, even children at our youngest age are capable of appropriately using the two words necessary for meaningfully employing our uncanny rating scale.



Figure II.6. Creepy-Weird Stimuli Used in Follow-Up Interview. Stimuli used for follow-up interview to address whether children appropriately understood the words “creepy” and “weird.”

Discussion

We provide three novel findings. First, the uncanny valley develops: younger children found the closely human-like and machine-like robot equally not very creepy, whereas older children found the closely human-like robot much creepier than the machine-like robot—similar to adults. Second, we identified the approximate age at which the uncanny valley emerges. Differences in feelings about the two focal robots emerged progressively over age, but it was not until middle childhood that children had a greater uncanny response to a closely human-like robot than a contrasting machine-like robot. Third, children’s perceptions of mind were correlated with this change in uncanny responses. For younger children, increasing perceptions of mind predicted *decreased* uncanniness. For older children, this association trended in the reverse direction, though not significantly so. Thus, from younger children to older children, the

correlation between mind and feelings of uncanniness increases from negative to trending positive, and the broader literature indicates that this correlation continues to increase and becomes positive for adults (K. Gray & Wegner, 2012). Of course, caution is needed in such cross-study child-adult comparisons (even though our videos for the Kaspars are exactly those used by K. Gray & Wegner, 2012). Regardless, we clearly demonstrate that feelings of uncanniness emerge and change over childhood and are associated with differing perceptions of robots' minds.

This research addresses important questions in psychology and robotics by providing initial evidence on the origins of the uncanny valley. One theory suggests that the uncanny valley is grounded in an innate mechanism (Steckenfinger & Ghazanfar, 2009), which means it should be present at an early age. Indeed, *a priori*, it is easy to imagine that young children could have responded to human-like robots as adults do—perceiving them as creepy. However, our results suggest that the uncanny valley emerges through development, and tracks changing understandings of mind. Our results clearly show that only children 9 years and older—those who have clear expectations about human and robot minds—feel unease towards very human-like robots, consistent with the second hypothesis outlined in the introduction that a robot is considered creepy when it violates our learned expectations of how a machine should look or behave.

The absence of the uncanny valley in younger children may reflect that they *expect* robots to have a myriad of mental abilities. In fact, other research on children's understanding of robots supports this speculation. Young children report that robots have perceptual abilities, like sight and touch: 3-year-olds claimed that a robot dog could see and be tickled (Jipson & Gelman, 2007). In conversations with parents, 3- to 5-year-olds also attributed biological, psychological,

and sensory abilities, in addition to features of artifacts, to a robot dog (Jipson, Gülgöz, & Gelman, 2016). As young children seemingly expect some robots to have mental abilities, the perception of a human-like mind may be a welcome familiarity for them. Indeed, our results are consistent with this explanation: young children found robots to be more pleasing (less uncanny) when they perceived the robots to have more mental abilities.

Older children, on the other hand, seemingly have different expectations about robot minds and expect robots to have reduced mental capacities. Five-year-olds are less likely to claim that a robot dog could think or feel happy compared to 3-year-olds (Jipson & Gelman, 2007) and are less likely to report that a robot has emotions, desires, or is capable of autonomous action (Mikropoulos et al., 2003). Fifteen-year-olds were also less likely to believe that the robot Robovie could have interests, experience emotions, or be a friend compared to 9- and 12-year-olds (Kahn et al., 2012). These changes in expectations with age have been linked to children's increasing experiences with and growing knowledge of technological devices (Bernstein & Crowley, 2008). Although the ages and robots varied across these studies, the general trend is clear: with age, children begin to deny psychological, emotional, social, and perceptual abilities to robots. For our older children, judgments of mind were no longer negatively associated with their ratings of uncanniness: attributions of mind stopped predicting a decrease in feelings of uncanniness. By hypothesis, older children in our sample may have evidenced an uncanny response to the human-like robot due to emerging changes in their expectations about the mental abilities of robots.

Our research was exploratory in the sense that, in advance of collecting the needed, relevant data, we had no firm prediction for which of various developmental patterns might

emerge. Although exploratory, our results clearly demonstrate that the uncanny valley effect develops with age, and that it occurs in tandem with judgments of mind.

Questions about the uncanny valley and children will become only more important over time as more robots are being made to interact and play with children. iPal, Jibo, and Zenbo, three Pixar-like robots, are designed to play games, answer questions, read stories, and watch children unsupervised (Glaser, 2016; Low, 2016; Wong, 2016). Nao, Ursus (a robotic bear), and Kaspar are all robots used to teach typically developing children, and those with motor disorders and autism spectrum disorders, a variety of skills including language (Movellan et al., 2009), physical exercises (Mejías et al., 2013), and social skills (Ricks & Colton, 2010). We should likely ensure that children, both typically developing and those with special needs, actually like these robots before extensively using them as companions or teachers.

While our research speaks to important avenues for future research, there are noteworthy potential limitations to our study. First, we acknowledge that our data are only correlational—unlike some work with adults (K. Gray & Wegner, 2012), there is no causal evidence for the link between understandings of mind and the uncanny valley in childhood. Future studies should therefore more explicitly test developments in thinking about minds (and machines) more generally with the development of the uncanny valley. This could involve both experiments and longitudinal studies that track children’s developing concepts of minds and machines (Gelman, 2003).

Second, for older children, judgments of uncanniness became dissociated with attributions of mind; for them, that link was no longer statistically significant. Although at first glance this result may seem problematic, a clear developmental picture emerges when these childhood data are coupled with data from adults in prior studies: the correlation between mind

and uncanniness increases over age from young (negative, i.e., increasing attributions of mind predict decreased feelings of uncanniness) to middle (zero) to older (trend positive) children and to adults (positive). On the other hand, it is also still possible that significant positive links between uncanniness and mind (increasing attributions of mind predict increased feelings of uncanniness) appear in middle to late childhood, in advance of adulthood. In order to be child-friendly, our scale for uncanniness had a restricted range (e.g., allowing three degrees of creepiness and feeling weird) in comparison with scales for adults, and it may be that a more nuanced scale would reveal similar effects to those with adults. A direction for future research would be to look further at participants aged 10 to adulthood but with a revised rating scale to assess the development of a positive association between mind and uncanniness in later childhood.

Third, our results do not speak to degrees of the uncanny valley, as we used a binary comparison for machine-like and very human-like robots (consistent with K. Gray & Wegner, 2012). The classic uncanny valley proposal is that liking of robots follows a non-linear curve (Mori et al., 2012) as in Figure II.1, and future research with children should explore the full range of its trajectory. Still, our study provides three initial data-points on this trajectory: a machine-like robot, an anthropomorphized robot, and a human-like robot. Although our study design (being a mix of within- and between-subjects) complicates the analysis of these three robot types somewhat, we can descriptively say that the anthropomorphic robot (Nao) was the least uncanny, followed by the machine-like robot, and finally the human-like robot as most uncanny.

It is moreover extremely likely that “humanness” is more than just a single dimension, thus plausibly robots could be human-like in several different multi-dimensional ways. Robots

could be closely human-like in face, limbs, behavior, language, and more. Future research should investigate which of these features of human-likeness is considered creepy and at what ages. And given our data, an important question would be which features are tied to mind? Future research examining a larger variety of robots across childhood is clearly needed.

Fourth, one might argue that our results do not speak to children's changing perceptions of robots but instead their developing understanding of words such as "weird," "feel," or "think." Our results with Nao, however, coupled with our additional data on young children's appropriate understanding of "weird" and "creepy" speak against this. Ample research also demonstrates that even preschoolers have appropriate understanding of the terms used in our questions eliciting children's attributions of mind, such as "think", "know", "feel" and "on purpose" (e.g., Bartsch & Wellman, 1995). Such findings make it difficult to argue that children have only shifted their understanding of the key words—creepy, weird, think, etc.—and are much more consistent with children appropriately using these terms to convey their developing conceptions of robots and robotic uncanniness and mind.

Understanding the development of the uncanny valley as an outgrowth of children's basic assumptions about robots coupled with increasing insights into minds provides a new perspective on this important phenomenon—it also suggests that one day the uncanny valley may disappear. As human-like robots become more commonplace and expand their abilities, children may come to expect that robots, although machines, *can* look surprisingly human, and *do* have minds, encompassing at least some human-like experiences. At which point, even highly human-like robots may be comfortingly familiar to children—even as they continue to unnerve today's adults.

Appendix II.A

Robot Beliefs Interview

Warm-up questions:

1. Do you like candy?
 - 1.1. “Do you like candy?”
 - 1.2. How much do you like candy? A little bit, a medium amount, or a lot?” or “How much do you not like candy? A little bit, a medium amount, or a lot?”
2. Do you like broccoli?
 - 2.1. “Do you like broccoli?”
 - 2.2. “How much do you like broccoli? A little bit, a medium amount, or a lot?” or “How much do you not like broccoli? A little bit, a medium amount, or a lot?”
3. Do you like carrots?
 - 3.1. “Do you like carrots?”
 - 3.2. “How much do you like carrots? A little bit, a medium amount, or a lot?” or “How much do you not like carrots? A little bit, a medium amount, or a lot?”

Interview questions:

4. Do you feel the robot is nice or creepy? +
 - 4.1. “Do you feel the robot is nice (thumbs up) or creepy (thumbs down)?”
 - 4.2. “How creepy do you feel it is? A little bit, a medium amount, or a lot

5. Does the robot make you feel weird or happy? +
 - 5.1. “Does the robot make you feel weird (thumbs down) or happy (thumbs up)?”
 - 5.2. “How weird does it make you feel? A little bit, a medium amount, or a lot?”
6. Would you want to play with the robot?
 - 6.1. “Would you want to play with the robot?”
 - 6.2. “How much would you want to play with it? A little bit, a medium amount, or a lot?”
7. Can the robot do things on purpose? * ++
 - 7.1. “Can the robot do things on purpose?”
 - 7.2. “How much can the robot act on purpose? A little bit, a medium amount, or a lot?”
8. When the robot moves, does it choose to move? +
 - 8.1. “When the robot moves, does it choose to move?”
 - 8.2. “How many things can the robot choose to do? A few things, a medium amount of things, or a lot of things?”
9. Does the robot think for itself? ++
 - 9.1. “Does the robot think for itself?”
 - 9.2. “How much does it think for itself? A little bit, a medium amount, or a lot?”
10. Some actions are bad, like hitting. And some actions are good, like helping. Does this robot know the difference between good and bad? +
 - 10.1. “Does this robot know the difference between good and bad?”
 - 10.2. “How much does it know the difference between good and bad? A little bit, a medium amount, or a lot?”
11. If I pinched the robot, would it feel pain? +
 - 11.1. “If I pinched the robot, would it feel pain?”

- 11.2. “How much can this robot feel pain? A little bit, a medium amount, or a lot?”
12. Does the robot have feelings, like happy and sad? * ++
- 12.1. “Does the robot have feelings, like happy and sad?”
- 12.2. “How much does the robot have feelings? A little bit, a medium amount, or a lot?”
13. If the robot saw a snake, would it feel scared? +
- 13.1. “If the robot saw a snake, would it feel scared?”
- 13.2. “How much can the robot feel scared? A little bit, a medium amount, or a lot?”
14. If the robot did not eat breakfast, would it feel hungry? +
- 14.1. “If the robot did not eat breakfast, would it feel hungry?”
- 14.2. “How much can the robot feel hungry? A little bit, a medium amount, or a lot?”
15. Is this robot like a human? *
- 15.1. “Is this robot like a human?”
- 15.2. “How much is the robot like a human? A little bit, a medium amount, or a lot?”
16. Does the robot know it’s a robot? * ++
- 16.1. “Does the robot know it’s a robot?”
- 16.2. “How much does it know it’s a robot? A little bit, a medium amount, or a lot?”

* item not included in final factor analysis due to cross loading or low factor loadings

+ item derived from Gray & Wegner, 2012 and/or H.M. Gray et al., 2007

++ item derived from Severson & Lemm, 2016

Chapter III

Robot Teachers for Children? Young Children Trust Robots Depending on their Perceived Accuracy and Agency

Abstract

Children acquire extensive knowledge from others. Today children receive information from not only people but also technological devices, and specifically social robots. Two studies assessed whether children appropriately trust and learn from technological informants. One-hundred-four 3-year-olds learned the names of novel objects from either a pair of social robots or inanimate machines where one robot or machine was previously shown to be accurate and the other inaccurate. Children trusted information from an accurate social robot over an inaccurate one and even more so when they perceived the robots as having psychological agency. Children did not learn from an inanimate, but accurate, machine however. Children can learn from technological devices (e.g., social robots) but trust their information more when the device appears to have mindful agency.

Keywords: trust in testimony, human-robot interaction, natural pedagogy, robots

Robot Teachers for Children? Young Children Trust Robots Depending on their Perceived Accuracy and Agency

Our world is increasingly filled with smart technological devices which inundate us and our children with information. We can uncover when dinosaurs went extinct, how far it is from the Earth to the Moon, and what the weather is like at the North Pole all from our laptops, phones, and now even from robots.

Robots are increasingly permeating our lives (National Robotics Initiative 2.0, 2017) and many are being designed to interact with, inform, and instruct children. Indeed, classrooms are increasingly implementing educational robots to teach children languages (Movellan et al., 2009), mathematics (Wei et al., 2011), science (Hashimoto et al., 2013), and, in some cases, even substitute for teachers (Han, 2010). This rapid expansion of robot “instructors” makes it increasingly important to investigate whether and when robots effectively transmit information to children.

Until recently, children’s knowledge exclusively came directly or indirectly from other humans. Children are adapted to learn about their world from the testimony of their parents, siblings, teachers, and peers (e.g., Csibra & Gergely, 2009). Children’s trust in testimony has classically been studied with an experimental paradigm where children learn the identities of novel objects from a pair of informants (Koenig, Clément, & Harris, 2004). From these studies, we know that children are capable of appropriately trusting a human informant, even at young ages. For example, children as young as 3 appropriately accept information from a consistently correct informant over a consistently incorrect one (Pasquini, Corriveau, Koenig, & Harris, 2007, p. 1216).

When children learn from other people, they monitor cues of psychological competency, like accuracy, but also who has access to the right kind of information (e.g., did she see the thing she is telling me about?); who is qualified (e.g., is she a knowledgeable adult or a naive child?); and who is confident (or uncertain) in their answers (e.g., did she say she *knows* it's an X or *thinks* it's an X?) (Koenig & Harris, 2007). Children similarly utilize social signals, like direct eye contact and contingent interactions, when determining whether they should accept information from informants (Csibra & Gergely, 2011). Arguably, children trust teachers that demonstrate psychological agency: the ability to think, make decisions, possess knowledge, and respond interactively (H. M. Gray et al., 2007).

Whereas children have relied on the testimony of humans to learn about their world for millennia, will they rely on the testimony of robots? Examining children's learning from the testimony of educational robots provides a unique means for investigating the scope of children's learning from smart technology and children's social learning more broadly. Specifically, social robots provide the opportunity to directly investigate social factors, like psychological agency, that impact children's trust in testimony. Unlike other forms of technology which have no agency or human instructors who have complete agency, the psychological agency of robots can be manipulated to assess its impact on social learning.

Indeed in a few tantalizing studies, when robots, like human instructors, demonstrate behaviors consistent with psychological agency, young children seem more willing to accept information from them. When learning about novel animals from two furry stuffed-animal-like robots, 3- to 5-year-olds were more likely to ask for information and agree with a robot that behaved contingently, turning to look at the child whenever he or she spoke, compared to one that did not (Breazeal et al., 2016). In another study, 4- to 6-year-olds were reported to be more

accurate when a robot used an interactive and contingent teaching style that required the children to perform tasks as a team with the robot (Okita, Ng-Thow-Hing, & Sarvadevabhatla, 2009). While intriguing, these prior studies have limitations (e.g., the furry robots were admittedly toy-like, rather than robotic) and have not directly tested the impact of psychological agency on children's trust in testimony.

In two studies, we address two key neglected questions in developmental science: (1) Do children appropriately trust (and mistrust) information from robots in the ways that they trust human informants? And (2) how does psychological agency impact young children's willingness to trust informants? Study 1 investigates whether young children appropriately trust social robots—socially-interactive robots with an abstractly human-like form. Study 2 provides contrasting baseline evidence that children will not simply trust or learn from just any inanimate technological object, specifically not two amorphous machines with no signs of psychological agency.

Study 1

Our paradigm was directly modeled on the classic one launched (first used by Koenig et al., 2004). Young children were taught the names of novel objects by two humanoid robots: one previously always accurate and the other always inaccurate. We focused on two possible outcomes: (1) children treat robots similarly to human instructors and appropriately trust only the accurate robot, or (2) children treat robots as distinct from human informants and indiscriminately trust either (or neither) robot.

To directly assess perceptions of psychological agency, children reported on their beliefs concerning the robots' mental abilities, including psychological agency, in a Robot Beliefs Interview.

Method

Participants. Fifty-nine 3-year-olds participated ($M = 3.6$ years; range = 3.0 to 4.0 years, 33 girls). Prior “trust in testimony” research clearly shows that 3-year-olds discriminate appropriately between accurate and inaccurate human informants (e.g., Pasquini et al., 2007). Power analyses indicate that $N = 33$ is sufficient to detect a moderate effect with .80 statistical power (Cohen, 1988). A moderate effect size was hypothesized because this effect size was observed in a comparable prior study (Pasquini et al., 2007). Eight other children were excluded for terminating participation ($N = 6$), being non-native English speakers (1), or parental interference (1). Children were primarily Caucasian (50) or bi-racial (8) by parent report.

Design. Children saw two robots (see Figure III.1) give contrasting testimony on the names of familiar and unfamiliar objects, judged which robot’s testimony they trusted, and then rated their perceptions of the robots’ capacities (e.g., “Can the robots think for themselves?”). The testimony procedure showed one robot correctly label a familiar object and a second robot incorrectly label it for four different familiar objects across four trials (the *Accuracy* trials). Following this came *Test* trials, where both the children and robots were presented with a novel, unfamiliar object and children were asked which robot they would ask for the object’s name (*Ask Question*). Then each robot gave a different name for the object and the child chose which name he or she thought was the actual name for the object (*Endorsement Question*). Children were also asked to judge which robot was not very good at naming the objects (*Accuracy Check Question*). A final additional set of questions measured children’s perceptions of the robots’ capacities—the Robot Beliefs Interview.



Figure III.1. Still Frame from Study 1. Still frame from Study 1 before the familiar or unfamiliar object is placed on the table.

The original task (Koenig et al., 2004) also used videoed informants. The robots we used, presented via video, and the questions in the Robot Beliefs Interview were validated for use with young children in a prior study (Brink et al., 2017).

Selective trust task. For each of four accuracy and four test trials, children watched a video of two Nao robots as in Figure III.1. The Nao robot is designed for use with children and has been used in prior research with children (Beran et al., 2013). Each trial began with a woman (see Figure III.1) placing an object between the two robots and asking each robot, “Can you tell me what this is called?” Each robot would then turn to look at the object, point at it, and name the object (e.g., “That’s a brush.”). (See Table III.1 for all robot responses across accuracy and test trials.)

Table III.1

Social Robot Responses for Each Trial

	Accurate Robot	Inaccurate Robot
Accuracy Trial 1	“That’s a brush.”	“That’s a plate.”
Accuracy Trial 2	“That’s a doll.”	“That’s a tree.”
Accuracy Trial 3	“That’s a ball.”	“That’s a cookie.”
Accuracy Trial 4	“That’s a bear.”	“That’s a towel.”
	Left Robot	Right Robot
Test Trial 1	“That’s a <i>modi</i> .”	“That’s a <i>toma</i> .”
Test Trial 2	“That’s a <i>gobi</i> .”	“That’s a <i>danu</i> .”
Test Trial 3	“That’s a <i>mogo</i> .”	“That’s a <i>nevi</i> .”
Test Trial 4	“That’s a <i>blicket</i> .”	“That’s a <i>terval</i> .”

Note: All robot responses to the question, “Can you tell me what this is called?” For half of the children, the accurate robot was on the left.

The two robots differed only in their accent colors, purple versus orange. The left-right position of each robot on the table was counterbalanced across participants. The order in which the robots were asked for the name of the object alternated across trials.

On *Accuracy* trials, children watched videos in which the same two robots named familiar objects. At the end of each trial, children answered a name-check question: e.g., “The orange robot said it’s a doll and the purple robot said it’s a tree. What do you think it’s called?” For the accuracy trials, the four objects were familiar (a hair brush, doll, ball, and teddy bear) and during test trials the objects were novel (see Figure III.2).

Familiar Objects				
Novel Objects				

Figure III.2. Familiar and Novel Objects Shown in Both Study 1 and Study 2. Unfamiliar objects were taken from the Novel Object, Unusual Name (NOUN) Database (Horst & Hout, 2014).

Following the accuracy trials, for test trials, children watched videos in which the same two robots named novel objects. Prior to the first test trial, children were shown a still frame of the video and asked an *Accuracy Check Question*: “One of these robots was not very good at answering these questions. Which robot was not very good at answering these questions?” (The trust in testimony literature often call this *Explicit Judgment*.)

At the start of each test trial, children reported which robot they would ask for the name of the novel object, the *Ask Question*: “I bet one of these robots can help us find out what it is called. Which robot do you want to ask, the orange robot or the purple robot?”

To end each test trial, each robot was asked for the name of the novel object and each responded with a different novel name (e.g., “That’s a *gobi*.” vs. “That’s a *danu*.”). At this point, children answered an *Endorsement Question*: e.g., “The orange robot said it’s a *gobi* and the purple robot said it’s a *danu*. What do you think it’s called? A *gobi* or a *danu*?”

In total, during the test trials, children were asked a series of nine focal questions: one Accuracy Check question, four Ask questions, and four Endorsement questions.

Robot Beliefs Interview (RBI). After the selective trust task, children were asked about their beliefs concerning the mental abilities of the robots and, specifically, whether they believed the robots had abilities related to psychological agency and perceptual experience. For *Psychological Agency*, children were asked whether the robots could think and make decisions: e.g., “When the robots move, do they choose to move?” and “Do the robots think for themselves?” For *Perceptual Experience*, they were asked if the robots could experience feelings: e.g., “If the robots saw a snake, would they feel scared?” The questions were formatted so as to result in a 4-point scale for each question coded as 0 (no), 1 (yes—a little bit), 2 (yes—a medium amount), and 3 (yes—a lot). We included perceptions of agency due to its apparent role in prior trust in testimony tasks with human instructors. Attributions of psychological agency have also been demonstrated to predict increased judgments of robot’s non-creepiness in children younger than 9 (Brink et al., 2017). We included perceptions of experience because they have been previously demonstrated to predict decreased affinity for robots in adults (K. Gray & Wegner, 2012) and provide additional information about robot’s minds beyond judgments of psychological agency alone (Brink et al., 2017). (For the complete interview, see Appendix III.A.)

Scoring the data. Six children were excluded from analysis because they failed to accurately answer the four name-check questions during the accuracy trials (1 failed 2, 5 failed 1) where they were asked to name the four familiar objects. For the remaining 53 children’s responses, the *Accuracy Check* question was scored correct when children identified the inaccurate robot as “not very good at answering these questions.” *Ask* questions were correct

when children reported that they would ask the accurate robot for the name of the novel object. *Endorsement* questions were correct when children selected the same novel word that the accurate robot reported for the novel object.

An aggregate for the robot's perceived psychological agency was calculated by averaging children's answers to three questions that assessed whether children believed the robots could think and make decisions: do the robots (1) "choose to move?", (2) "think for themselves?", (3) "know the difference between good and bad?" An aggregate for the robot's perceived perceptual experience was calculated by averaging children's responses to three questions that assessed whether children believe robots could feel: would the robots (1) "feel pain?", (2) "feel scared?", (3) "feel hungry?" The use of these six questions and their aggregation in this fashion was validated by Exploratory Factor Analysis of responses for a larger set of items from 239 children aged 3 to 18 years (Brink et al., 2017).

Results

Performance on test trials. As shown in the left-hand panel of Figure III.3, children performed above chance for all three test question types: (1) for the *Accuracy Check* question, $M = .96$, $t(52) = 17.49$, $p < .001$, 95% CI [.91, 1.02]; (2) for *Ask* (proportion correct for which robot would they ask for the name of the novel object), $M = .63$, $t(52) = 3.44$, $p = .001$, 95% CI [.55, .70]; and (3) for *Endorsement* (proportion correct for what they thought the novel object was named), $M = .68$, $t(51) = 4.11$, $p < .001$, 95% CI [.59, .77]. Children's appropriate discrimination of information from accurate and inaccurate robots is highly similar to comparable performance for 3-year-olds with human informants. In Pasquini et al. (2007), children appropriately asked and endorsed the accurate human informant in 70% and 67% of the trials respectively.

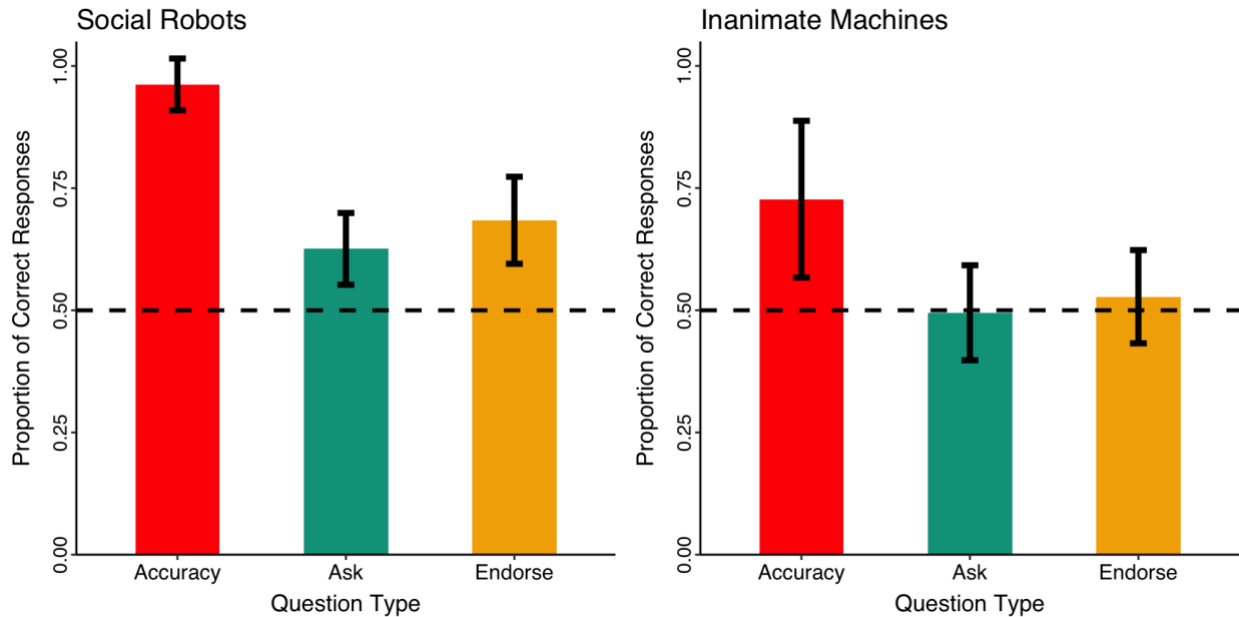


Figure III.3. Proportion of Correct Responses to Accuracy, Ask, and Endorse Questions. The left panel shows the proportion of correct responses to Explicit Judgment, Ask, and Endorse questions for Study 1 (social robots). The dashed line represents performance at chance. Children were above chance for all question types in Study 1. The right panel shows responses for Study 2 (inanimate machines). Children were above chance for only the Accuracy Check question in Study 2. The error bars represent 95% confidence intervals.

Beliefs about robots as predictors of performance. Answers to the three *Agency* questions (e.g., “When the robot moves, does it choose to move?”) were calculated as the average score of all three questions. Similarly, answers to the three *Experience* questions (e.g., “If the robot saw a snake, would it feel scared?”) were calculated as the average score of all three questions. Children’s scores for psychological Agency ranged from 0 to 3 ($M = 1.82$, $SD = .91$) and were significantly different from 0, $t(47) = 13.82$, $p < .001$. Children’s scores for Experience ranged from 0 to 3 ($M = 1.99$, $SD = 1.02$), and were significantly different from 0, $t(47) = 13.52$, $p < .001$.

We used these and other factors to predict children’s performance on the selective trust task as measured by their Total Correct performance—the proportion of children’s correct

responses to all eight Ask and Endorsement questions. Thus, children could have scores ranging from 0 (responded incorrectly to all questions they answered) to 1.0 (responded correctly for all answered test questions). Children's scores ranged from 0 to 1 ($M = 0.65$, $SD = 0.25$). We then assessed whether children's answers on the Robot Beliefs Interview predicted total composite scores.

In a regression analysis with Agency, Experience, Age, Sex, and Order (4 orders combining robot left-right placement and which robot was accurate) as predictors of Total Correct performance, Agency significantly predicted performance, $\beta = .29$, $t(46) = 2.08$, $p = .044$, $R^2 = .425$. As attributions of the robots' psychological agency increased, appropriate performance on the selective trust task increased. Age also predicted performance on the task, $\beta = .30$, $t(46) = 2.37$, $p = .023$. Older children were better at asking and agreeing with the accurate robot. One variation of the presentation order—children performed better when the purple robot was accurate and standing on the left side of the table—also predicted better performance on the task, $\beta = 1.12$, $t(46) = 3.43$, $p = .001$. We do not discuss this last result further; it is likely a spurious result as there is no substantive reason why children should perform better depending on the particular side that the robot is standing and on and for only the purple robot.

Discussion

Children trusted the accurate robot over the inaccurate robot: They were more likely to ask for and agree with information provided by the accurate robot. Thus, children as young as 3 are willing and able to appropriately trust accurate robots and mistrust inaccurate robots. They also learn from the accurate robot in the minimal sense of repeating the object name provided by the accurate robot when asked what they think a novel object is called.

Further, as children's perceptions of psychological agency for the two robots increased, they were more likely to appropriately trust the accurate robot. Generally, these young children attributed agency and experience to these robots, but there was also variation in how strongly they did so. Children who more strongly attributed the ability to think and make decisions to the robots were more likely to ask for and endorse information from the accurate robot. This finding supports the hypothesis that children are increasingly likely to treat social robots similarly to human teachers when those robots engender perceptions of agency.

Study 2

Study 2 experimentally probed the impact of psychological agency on children's trust and learning from machine informants. In Study 2, the social robots were replaced with simple inanimate machines, one purple and one orange. Compared to the social robots, the machines provided no obvious cues of psychological agency: they were amorphous blob shapes without faces or limbs and did not respond contingently to the woman presenting the familiar objects. Instead, the woman signaled a person off-camera whose hand reached on-screen to pull a string that activated an audio recording naming the objects. If children only monitor the accuracy of entities that demonstrate psychological agency, they should be less likely to monitor the accuracy of these two machines and neither selectively ask nor endorse either one.

This design also addressed an alternative low-level interpretation of our results from Study 1. The accuracy trials may have set up a simple positive bias for one robot over the other, such that children simply asked for and endorsed information based on that bias rather than an explicit recognition of accuracy. This is a non-trivial hypothesis worth addressing because at one level our "informants" in Study 1 are simply devices whose different colors might act at the level of stimulus cues (possibly resulting in cued discrimination conditioning). If children were merely

associating the color cues for each robot in Study 1 with correct and incorrect naming, then they could be expected to perform similarly in Study 2, appropriately asking for and endorsing information from the accurate machine.

Method

Participants. Forty-five 3-year-old children identical in age ($M = 3.5$ years; range = 3.0 to 4.0 years, 21 girls) and background to those in Study 1 participated. Five additional children were excluded for terminating participation ($N = 2$), being non-native English speakers (1), or parental interference (2). Children were primarily Caucasian (33) or bi-racial (9) by parent report. Power analyses indicate that $N = 33$ is sufficient to detect a moderate effect size with .80 statistical power (Cohen, 1988) just as in Study 1.

Design. The design was identical to that of Study 1, except the robots were replaced with two blob-like “machines” (depicted in Figure III.4). Children saw the two machines provide contrasting testimony, judged which machines’ testimony to trust, and then rated their perceptions of the machines’ capacities (e.g., “Can the machines think for themselves?”).

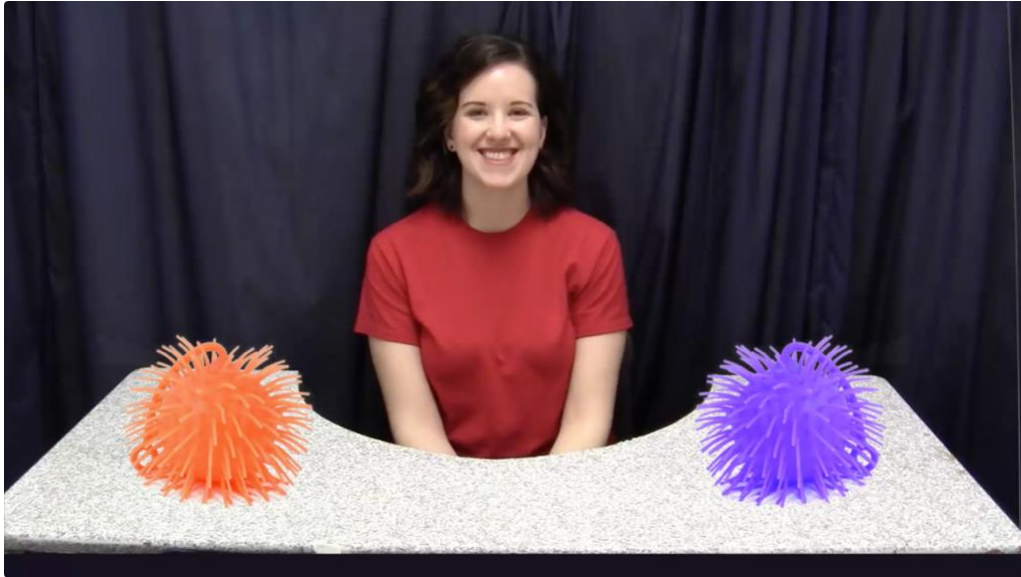


Figure III.4. Still Frame from Study 2. Still frame from Study 2 investigating whether children appropriately trust (and mistrust) two inanimate machines with minimal signs of psychological agency.

Selective trust task. The procedure was identical to that of Study 1 with some minor changes: (1) informants were referred to as “machines” rather than “robots” and (2) the woman addressed someone offstage rather than the machines themselves, after which a hand came in to pull a string activating each machine’s “voice.” The machines spoke with the exact same robotic voice as in Study 1. (For all machine responses across accuracy and test trials see Table 3.2.)

Table III.2

Inanimate Machine Responses for Each Trial

	Accurate Machine	Inaccurate Machine
Accuracy Trial 1	“A brush.”	“A plate.”
Accuracy Trial 2	“A doll.”	“A tree.”
Accuracy Trial 3	“A ball.”	“A cookie.”
Accuracy Trial 4	“A bear.”	“A towel.”
	Left Machine	Right Machine
Test Trial 1	“A <i>modi</i> .”	“A <i>toma</i> .”
Test Trial 2	“A <i>gobi</i> .”	“A <i>danu</i> .”
Test Trial 3	“A <i>mogo</i> .”	“A <i>nevi</i> .”
Test Trial 4	“A <i>blicket</i> .”	“A <i>terval</i> .”

Note: All machine responses. For half of the children, the accurate machine was on the left.

Machine beliefs interview. After the selective trust task, children were asked about their beliefs concerning the mental abilities of the two machines using an interview nearly identical to the interview in Study 1 except that references to “robots” were replaced with references to “machines.”

Scoring the data. Twelve children were excluded from analysis because they failed to accurately answer all 4 name-check questions (2 failed 3, 6 failed 2, 4 failed 1). For the remaining 33 (17 girls) children’s responses, proportion of correct responses for Accuracy Check, Ask, and Endorsement questions were calculated. Aggregates for children’s judgments and answers were formed just as in Study 1.

Results

Performance on test trials. As shown in the right-hand panel of Figure III.3, children were above chance for only *Accuracy Check* questions, $M = .73$, $t(32) = 2.89$, $p = .007$, 95% CI [.57, .89]. However, children performed at chance for *Ask*, $M = .49$, $t(32) = -.11$, $p = .916$, 95% CI [.40, .59]); and for *Endorsement* questions, $M = .53$, $t(32) = .59$, $p = .557$, 95% CI [.43, .62].

In a follow-up analysis, we examined *Ask* and *Endorse* responses for only those children who correctly identified the accurate versus inaccurate machine on their *Accuracy Check* question. This more closely parallels the analyses in Study 1, where children answered that question with 96% accuracy. These 24 children were still at chance for *Ask*, $M = .48$, $t(23) = -.29$, $p = .773$, 95% CI [.36, .61]; and *Endorsement* questions, $M = .54$, $t(23) = -.75$, $p = 0.461$, 95% CI [.43, .64].

In Study 1, children's strength of ascribing various "mind" properties, and specifically psychological agency, predicted their willingness to trust and learn from the robots. In Study 2, although there was sufficient variance, with some children attributing more and some less psychological agency to the machines ($M = 1.69$, $SD = .96$), children's strength of ascribing psychological agency failed to predict their willingness to trust and learn from the machines, $r(28) = .211$, $p = .3$. This is likely because, in Study 2, on average children chose randomly between the two machines regarding who they should *Ask* or *Endorse*. Thus, even when they knew which machine had been accurate, they trusted neither one's answers regarding object names.

Discussion

Study 2, in comparison with Study 1, investigated whether perceived psychological agency causes children to appropriately trust accurate machine-like informants. When cues of agency for the two machines were minimized, children did not utilize the accuracy of these inanimate machines to decide which to trust. Children's responses thus support the hypothesis that children especially value informants with psychological agency. Arguably, for children, it is important that an informant not only be accurate but demonstrate the ability to think and make decisions.

These results also make unlikely that low-level associations alone account for children's performance in Study 1. Children recognized which machine was "not very good at saying toy names," demonstrating sufficient attention to and interest in the crucial distinction between the devices. However, children were at chance when "trusting" either machine.

General Discussion

We found that: (1) Young children appropriately learn from and trust (and mistrust) information from robots similarly to the way they trust humans. For social robots, as for human informants, trust is granted and withheld based on informant accuracy. (2) Psychological agency impacts young children's willingness to trust informants. Children appropriately trusted the accurate robot to the extent that they attributed more psychological agency to the robots (Study 1). Further, when agency cues were reduced, children were indifferent to information from either an accurate or inaccurate machine (Study 2).

Our findings extend the broader literature on children's trust in testimony by further highlighting the importance of an informant's mental capacities when determining whether they can be trusted. Children sought out and agreed with information from an informant perceived to think and make decisions. Previous research demonstrates that children attend to signals about their human instructors' mental abilities—accuracy, confidence, expertise, and access to information—which can also be considered a by-product of the foundational ability to think and make decisions. It may be that only when informants have psychological agency in these ways, that young children begin to assess additional factors: is my informant accurate, confident, knowledgeable? At the least, psychological agency seems to be a critical factor in determining whether children trust the testimony of social robots and may well be important more generally.

Relatedly, our findings have implications for the use of educational robots. Robots have been working in classrooms, daycares, clinics, and hospitals where they are increasingly expected to convey and pass on knowledge to children, including language skills (Movellan et al., 2009), mathematics (Wei et al., 2011), and science (Hashimoto et al., 2013). As we demonstrate, the agent-like qualities of robots have the potential to impact the quality of children's learning experiences with these devices. By implication, designers should consider the impact of agent-like cues (e.g., contingent behaviors or abstractly human-like features) when building educational robots.

Educational devices are not limited to the agentic robots or purely inanimate devices we utilized. Educational tools can range in agent-like cues from disembodied intelligent assistants, like Apple's Siri, to physical humanoid robots like Nao. Likewise, they can vary in the quality of their contingent responses, social interactions, and abstractly human-like appearances. We examined only one device with numerous agent-like cues (i.e., contingent responses, referential speech, and abstractly human-like features) versus one with no such cues whatsoever. In future research, it will be useful to examine the effectiveness of devices which vary in the quantity and quality of features that encourage attributions of agency.

Robots are being used not only with typically-developing children, but in educational endeavors for children with special needs. Several robots are helping children with ASD engage in social interaction through imitation games, turn-taking, and conversation (Ricks & Colton, 2010). Within this literature, the impact of agency on the effectiveness of robot instructors for ASD children is debated. Although more attractive to children with ASD (Robins, Dautenhahn, & Dubowski, 2006), non-human-like robot instructors are hypothesized to prevent children with ASD from transferring their learned social skills to real people (Ricks & Colton, 2010).

Additional research is needed to ascertain the importance of attributions of agency for special populations.

Future research could also delve deeper into learning. As one example, do young children evidence word learning from a robot after a delay? This too, clearly, has educational implications.

In advance of such research, we demonstrate that young children learned from and appropriately trusted (and mistrusted) information from a humanoid technological device and that an important factor in this trust was the extent to which they perceived the device to be able to think and make decisions. These findings set the stage for numerous intriguing and important future studies.

Appendix III.A

Reduced Robot Beliefs Interview

1. Do you like candy? (Warm-Up Question)
 - 1.1. “Do you like candy?”
 - 1.2. “How much do you like candy? A little bit, a medium amount, or a lot?” or “How much do you not like candy? A little bit, a medium amount, or a lot?”
2. When these robots move, do they choose to move?
 - 2.1. “When these robots move, do they choose to move?”
 - 2.2. “How many things can the robots choose to do? A few things, a medium amount of things, or a lot of things?”
3. Do the robots think for themselves?
 - 3.1. “Do the robots think for themselves?”
 - 3.2. “How much do they think for themselves? A little bit, a medium amount, or a lot?”
4. Do these robots know the difference between good and bad?
 - 4.1. “Some actions are bad, like hitting. And some actions are good, like helping. Do these robots know the difference between good and bad?”
 - 4.2. “How much do they know the difference between good and bad? A little bit, a medium amount, or a lot?”
5. If I pinched the robots, would they feel pain?
 - 5.1. “If I pinched the robots, would they feel pain?”
 - 5.2. “How much can these robots feel pain? A little bit, a medium amount, or a lot?”

6. If the robots saw a snake, would they feel scared?
 - 6.1. "If the robots saw a snake, would they feel scared?"
 - 6.2. "How much can the robots feel scared? A little bit, a medium amount, or a lot?"
7. If the robots did not eat breakfast, would they feel hungry?
 - 7.1. "If the robots did not eat breakfast, would they feel hungry?"
 - 7.2. "How much can the robots feel hungry? A little bit, a medium amount, or a lot?"

Chapter IV

A Theory of the Robot Mind: Developing Beliefs about the Minds of Social Robots

Abstract

This chapter investigates two research questions more broadly: (1) which features of robots encourage attributions of mind? And (2) and how do these features encourage feelings of uncanniness? Utilizing the Robot Beliefs Interview, 473 adults and 120 children (3- to 18-years-old) rated their feelings toward and beliefs about ten distinct robots. These ten robots varied on several dimensions theoretically linked to attributions of mind and uncanniness: purpose, attractiveness, and resemblance to humans, animals, or fictional characters. By investigating adults' and children's responses to a wider variety of robots, this study informs an understanding of the uncanny valley, the attribution of mind to machines, and the factors that produce these outcomes.

Keywords: uncanny valley, theory of mind, social cognition

A Theory of the Robot Mind: Developing Beliefs about the Minds of Social Robots

We live in a world increasingly filled with smart technology—laptops, tablets, smart phones, autonomous cars, and now robots. And, every day, these devices are behaving more and more like people, as if they have minds of their own. Robots are special and unique devices that can simultaneously share similarities with artifacts, animals, and even humans. They, like other artifacts, are designed and built by humans. Yet, unlike more common artifacts like books, bikes, beds, and balls, social robots can look, behave, and at times even “think” like humans or animals. Indeed, adults and children often get the impression, after watching or interacting with robots, that they do in fact think like we do.

Utilizing survey methods, several prior studies show that adults and children alike will attribute mind to robots to varying degrees (Brink et al., 2017; H. M. Gray et al., 2007; Kahn et al., 2012; Weisman, Dweck, & Markman, 2017). Moreover, attributions of a human-like mind to social robots are an important factor in determining feelings toward robots (Brink et al., 2017; K. Gray & Wegner, 2012) and willingness to accept information from them (Brink & Wellman, submitted). Attributions of mind to robots occur across ages and impact our interactions with robots, but what factors cause us to attribute minds to robots?

One possibility is that a robot’s appearance and design affect our willingness to attribute a mind to it. To create precise experimental contrasts, previous research has focused on children’s and adults’ responses to a limited set of robots that represent only a small number of the broad range of existing robots. Robots, however, are not limited to simply human-like, machine-like, or humanoid robots. Instead, robots can vary substantially in their appearances and designs, resembling any of a number of living creatures or appearing as entirely new and unique entities. Due to the existence of drastically different types of robots, it is important to explore

how children and adults react to the appearance and behaviors of a broad number of robots. What physical and behavioral features of robots contribute to attributions of mind? And, in turn, how do they impact uncanniness? The studies presented here more carefully explore the specific features of a broad range of robots and how those features may group together, encourage attributions of mind, and influence feelings of uncanniness (primarily throughout adulthood but also in childhood).

One theory proposes that human-likeness leads to attributions of mind and in turn uncanniness responses to robots. The uncanny valley (see Chapter II) has historically been argued to result from levels of human-likeness in the robot itself. Human-likeness, however, has not been consistently defined across studies. Often, a robot's human-likeness is defined a priori by an experimenter (e.g., K. Gray & Wegner, 2012; Mori et al., 2012) or by asking adult participants to directly rate human-likeness (MacDorman, 2006). These ratings of human-likeness have not consistently predicted uncanniness across studies: whereas some "human-like" robots have been empirically confirmed to be creepier than machine-like robots, not all "human-like" robots produce the same uncanny responses (Hanson, 2006; MacDorman, 2006). Thus, the use of these ratings of human-likeness have yet to fully capture the non-linear shape of the uncanny valley originally proposed by Mori (see Mori et al., 2012). Either these ratings have only captured a segment of the shape of the uncanny valley, or, worse yet, they contradict the theoretical shape entirely (Hanson, 2006; Hanson et al., 2005; MacDorman, 2006).

One useful perspective that may explain these failures to empirically validate the full non-linear shape of the uncanny valley is that human-likeness, though historically treated as unidimensional across these studies, is in fact multidimensional. Plausibly, robots could be human-like in several different multi-dimensional ways. Kaspar (Chapter 1) is increasingly

human-like in mimicking human facial features. But Nao (Chapters 1 and 2) is human-like in a much less realistic, anthropomorphized way—e.g., it has two arms and two legs (not four legs), hand- and foot-like extremities (not claws or wings), and a human head-like shape perched vertically on its body (not a forward thrusting head like a dog). Robots could be closely human-like in limbs, face, behavior, language, and more. Figure IV.1 shows one sample of various robots that vary on some of these categories of features. These static depictions alone show that, while human-likeness has often been treated as a unidimensional measure, there is ample reason to believe that human-likeness, instead, consists of multiple components. Potentially, because human-likeness, as of yet, has not been clearly operationalized, the full non-linear shape of the uncanny valley itself has never been empirically verified.

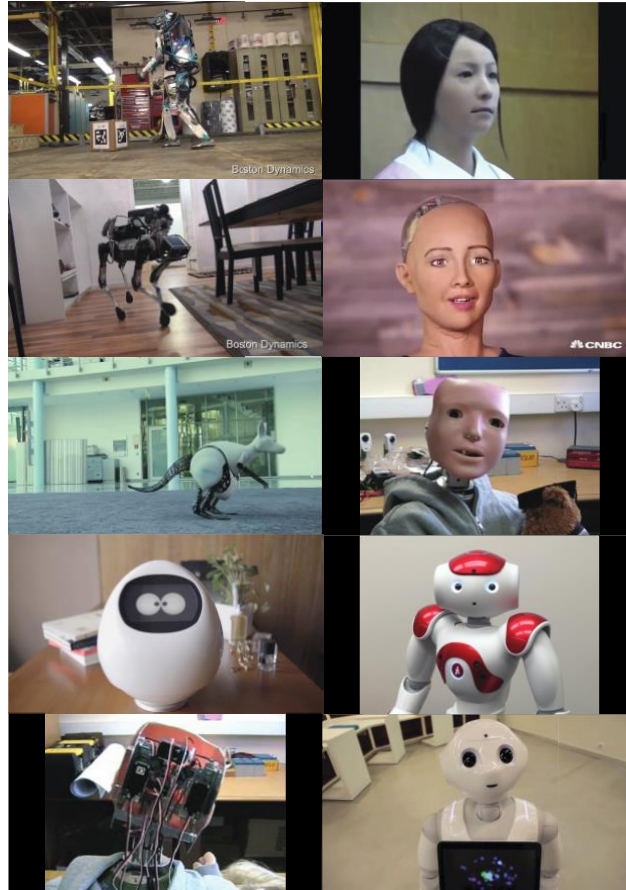


Figure IV.1. Still Images of 10 Robots Presented in Study 1. Still images of 10 robots presented in Study 1 of Chapter 4. Reading left to right and down the images the robots are: Atlas, Actroid, Spot, Sofia, Festo, Kaspar front, Tapia, Nao, Kaspar back, and Pepper. As in Study 1 of Brink et al. (2017) participants only ever saw Kaspar front or Kaspar back (no individual participant ever saw both). That way, participants could never deduce these were alternative views of the same robot. The video clip for each robot can be found in the online supplemental materials. (Note: These videos will be available for viewing at the oral defense).

Alternatively, attributions of mind to a robot and feelings of uncanniness may not be related to the human-likeness of that robot at all. Indeed, a variety of visual features have been linked to attributions of agency. Simple static perceptual features, like eyes, can inspire attributions of agency: e.g., infants attribute moral agency to the actions of objects with googly eyes but not to identical objects without them (Hamlin, Wynn, & Bloom, 2010). Dynamic features can also encourage attributions of psychological agency. For instance, adults and even

infants attribute intentions to the movements of blocks, triangles, and blobs (Heider & Simmel, 1944; Luo & Baillargeon, 2005; Shimizu & Johnson, 2004). Some such features can best be conveyed in the interactions of objects, in which objects give the impression that they are chasing or avoiding one another (see research with adults by Michotte (1963/2017) and Scholl & Tremoulet (2000); and with infants by Schlottman, Ray & Surian (2012)). Behaviorally, infants also react differently to objects that interact contingently with the infant (S. C. Johnson, 2000). In the set of studies featured in this chapter, I present dynamic but not interactive features.

Further, attributions of mind and uncanniness responses may also or alternatively be the result of factors unique to the individual viewing the robot and not the robots themselves. Epley and colleagues (2007) argue that dispositional (e.g., an individual's need for cognition), situational (e.g., perceived similarity to the nonhuman agent), cultural (e.g., experience, norms, and ideologies), and developmental (e.g., acquisition of alternate theories) factors impact whether we attribute mental abilities to nonhuman agents. The appearance of the robot may be less important than, or work in conjunction with, a person's unique experiences and background to impact their attributions of mind and uncanniness responses to robots.

Across two studies, I explore whether combinations of attributes within a robot or whether the individuals themselves are the deciding factor on whether we attribute minds to robots and experience feelings of uncanniness. These studies address four focal research questions: (1) ***Mental attributions to robots***: How do adults and children attribute mental abilities to real-world robots across a much wider, varied sample of robots (compared to previous studies)? (2) ***Characteristics of robots***: Do the characteristics of robots, like appearance and behavior, predict the types of mental attributions that adults make and the strength of those mental attributions? And if so, how? (3) ***Individual differences and consistencies***: Do adults and

children change how they attribute mental attributes to robots depending on the type of robot, or do individuals consistently apply the same mental attributions to robots regardless of robot? (4)

Robot uncanniness: How do mental attributions impact adults' and children's uncanniness responses to a wide variety of real-world robots?

In the initial study, I explore these questions in adults to first define the general pattern of behavior. And, in a second study, I begin to explore the answers to these questions in children. Initial incomplete data from over 100 children (Study 2) will be described and compared to the adult data (Study 1). Do children show the same pattern as adults? If not, at what stage do they?

Study 1: Adults

Method

Participants. 473 (190 women, 3 another gender identity, 5 refused to answer) adult participants, mean age 36.4 (20, 75), were recruited using Amazon's MTurk. All participants: (1) had an approval rating greater than or equal to 95 (out of 100), (2) had completed at least 50 approved tasks, (3) were located in the United States, and (4) were over 18 years old.

Participants were compensated \$0.50 for completing the survey. Eleven were excluded for failing to report a birth year, 36 participants were excluded for completing the survey in less than 2 minutes or taking more than 15 minutes.

Task and design overview. Participants rated five videos of real-life robots—randomly selected from a set of 10 videos of real-life robots—and one video of a person. For each robot, participants answered 10 questions concerning their feelings about the robot and their beliefs about the mental abilities of that robot. With five videos of robots and one video of a person lasting 8 secs each and approximately 50 questions, the duration of the study averaged 4 minutes and 46 seconds.

Videos. Each video presented a robot or person for approximately 8 seconds (range 7-9 seconds). Within each video, the robot or a person was shown either talking to someone off camera or locomoting around a room. All videos were presented without sound. Still images of each robot can be seen in Figure IV.1. The 10 robots sampled for this study were chosen because (1) they have video clips available online, (2) many are commercially available, and (3) they intuitively range over a variety of distinct features that plausibly impact attributions of mind and feelings of uncanniness (e.g., presence or absence of human-like features—eyes, mouth, speech, skin-like covering, intentional action—to varying degrees).

Adult Robot Beliefs Interview (RBI). A subset of eight interview questions were taken from Brink et al.'s (2017) Robot Beliefs Interview (RBI; originally designed for children) relevant to the variables of interest: feelings of *Uncanniness* and attributions of *Psychological Agency* and *Perceptual Experience*. Questions were modified for adults and to fit an online survey format.

Uncanniness. For each robot video, adults were assessed on their uncanny response via two questions gauging the extent to which they felt the robot was creepy or unsettling: (1) “Do you feel the robot is creepy?” and (2) “Does the robot make you feel weird?” Paralleling the methods established in Brink et al. (2017) the response format resulted in a 4-point scale for each question coded as 0 (No), 1 (Yes-a little bit), 2 (Yes-a medium amount), and 3 (Yes-a lot).

Mental attributions. Adults also answered seven questions assessing the robots' mental capacities for *Psychological Agency* and *Perceptual Experience*. Agency was captured with three items: “does the robot” (1) “purposefully choose to move?”, (2) “think for itself?”, and (3) “know the difference between right and wrong?” Experience was captured by three items: “would the robot” (1) “feel pain?”, (2) “feel scared?”, (3) “feel hungry?” (See Appendix IV.A for

complete adult interview.). An additional exploratory question asked, “Does this robot have feelings, like happy and sad?” Each question resulted in a 4-point scale coded as 0 (No), 1 (Yes-a little bit), 2 (Yes-a medium amount), and 3 (Yes-a lot).

Robot validity check. We included one additional item to check whether participants were properly aware that the stimuli were either a human or robot. Immediately after watching each video, adults were asked to report whether they believed the agent in the video was a robot or a person. After responding, adults were given the correct identification of the agent in the video.

Results

Research questions. Analyses addressed our four focal research questions: (1) ***Mental attributions to robots:*** How do adults attribute mental abilities to real-world robots across a much wider, varied sample of robots (compared to previous studies)? (2) ***Characteristics of robots:*** Do the characteristics of robots, like appearance and behavior, predict the types of mental attributions that adults make and the strength of those mental attributions? And if so, how? (3) ***Individual differences and consistencies:*** Do adults change how they attribute mental attributes to robots depending on the type of robot, or do individuals consistently apply the same mental attributions to robots regardless of robot? (4) ***Robot uncanniness:*** How do mental attributions impact adults’ uncanniness responses to a wide variety of real-world robots?

Data analysis plan. *Mental attributions to robots.* Using both Confirmatory Factor Analysis (CFA) and Principal Component Analysis (PCA), we identified the types of mental attributes that adults applied to our large sample of robots and compared these mental attributions to those identified in previous studies with far fewer robots (Brink et al., 2017; H. M.

Gray et al., 2007). We then measured and examined the strength of these mental attributions across robots and participants.

Characteristics of robots. Using clustering techniques, we determined whether types and strengths of mental attributions typically co-occurred with specific robot appearances or behaviors.

Individual differences and consistencies. Using the same clustering techniques, we also investigated whether participants were consistent in their mental attributions to robots. We assessed whether adults uniformly applied the same types and strengths of mental attributions to all robots or whether participants applied varied mental attributions across robots.

Robot uncanniness. We used regression modeling to investigate whether types and strengths of mental attributions predicted uncanniness to determine the factors that best account for raters' feelings about the attractiveness or uncanniness of robotic agents.

Imputation. We applied missing value imputation to the data set to increase statistical power. Because we randomly assigned adults to view five of 10 videos, we concluded that the missing data for each participant for the five remaining videos was missing completely at random. We utilized nonparametric missing value imputation using Random Forest (Stekhoven, 2013). This method is appropriate for mixed-type datasets with many variables that encompass both continuous and categorical data including complex interactions and nonlinear relations. Its out-of-bag (OOB) imputation error estimate for our data was 12.95%.

All analyses were performed both with a fully imputed data set and the original data set. However, only the results from the original data set are described here. We describe only the original data set because those data adhere most closely to participants' actual per robot ratings, although this, of course, results in fewer total data points for subsequent analyses. Moreover, this

analysis method provides a better comparison to the data from children (see Study 2 below). The original (non-imputed) dataset alone is large, it provides 2079 separate sets of ratings on the 10 robots (with each robot rated on 10 different perceptions).

Robot validity check. Seven of the 10 robots were identified as robots by 100% of the participants. All 10 robots were identified as robots by at least 95% of participants. Less than 2% of participants identified the video of the person as a robot. The high accuracy of our participants indicates both that the robots were easily identifiable as robots and that participants were focused on the task.

Mental attributions to robots. *Confirmatory Factor Analysis (CFA)*. We performed a CFA to determine whether adults attribute mental abilities to a broader range of robots similarly to adults in previous studies (Brink et al., 2017; H. M. Gray et al., 2007). A diagram of the CFA model can be seen in Figure IV.2.

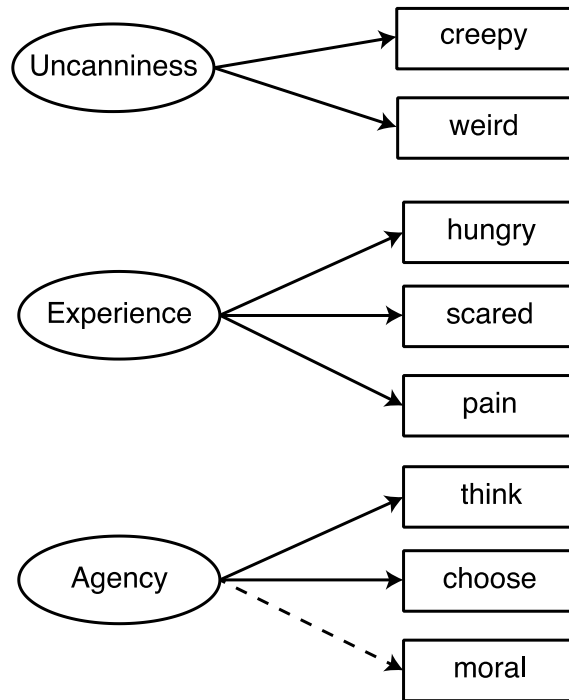


Figure IV.2. A Diagram of the CFA Model for Studies 1 and 2. The *moral* item was not a good fit of the data for Study 1, but *moral* did fit with *think* and *choose* in Study 2 with children; this is represented by the dashed line.

An initial CFA model, that included all items from the RBI, was not a good fit of the data. When the *moral* survey item was excluded from the analyses, however, the fit indices supported the modified CFA model. Four commonly used model fit indices were all within their respective cutoff ranges (RMSEA = .046 < 0.05; CFI = 0.994 > 0.95; TLI = 0.989 > 0.95; SRMR = .021 < .08) (Hu & Bentler, 1999). Nearly all factor loadings met or approached the threshold for good indicators of .7 (Kline, 2011). For Agency, its two indicators had factor loadings of 0.561 (*choose*) and 1.076 (*think*). The three indicators for Experience were within 0.851-0.911. For Uncanniness, the two indicators were 0.789 (*creepy*) and 1.088 (*weird*).

From the CFA analysis, we created three aggregates: two types of mental attribution aggregates—Psychological Agency and Perceptual Experience—and an Uncanniness aggregate.

Each aggregate was calculated by averaging participants' responses for each aggregate's respective items (as shown in Figure IV.2).

Principal Component Analysis (PCA). To increase power for subsequent analyses, we utilized a second data reduction method, PCA, which retains all interview items. PCA calculates its comparable components by including and weighting every survey item. CFA, on the other hand, calculates its factors with weights of either 0 or 1, effectively excluding those items with weight 0 and thus utilizing a reduced set of the original data. Whereas principal components are less easily interpretable than factors, they retain more information about participants' responses for predictive analyses. Our PCA included information from two additional items that were not included in the CFA: *feel* and *moral* items. Because we ultimately wanted to predict uncanniness, all uncanniness items were excluded from the PCA analysis.

The two components extracted from the PCA explained 78.8% of the variance in our data. The first principal component (PC1) weighted most heavily the items *feel*, *scared*, *hungry*, *pain*, and *moral*. This first principal component explained 60.7% of the variance. The second principal component (PC2) placed more weight on *think* and *choose* and explained 18.2% of the variance. In these ways the PCA ultimately produced substantively similar, but quantitatively distinct types of mental attributions compared to the CFA: PCA of participants' ratings similarly divided the survey items into Agency (PC2) and Experience (PC1) compared to CFA but calculated them with different item weights and took into account ratings of *feel* and *moral*.

The principal components calculated by the PCA were utilized in subsequent analyses and also compared to the results for the CFA.

Clustering. The two aggregates, factors (from CFA) and principal components (from PCA), of the types of mental attributions—Agency and Experience—were analyzed using

unsupervised statistical learning methods (i.e., K-means clustering, hierarchical clustering). These techniques can empirically identify distinct categories of responses as defined by the unique combinations of the strength and types of mental attributions. For example, clustering analyses may cluster responses into categories like “high agency, high experience (HAHE)”; “high agency, low experience (HALE)”; “low agency, high agency (LAHE)”, or “high agency, low experience (HALE).”

For our 426 participants, we obtained 2079 sets of ratings across 10 robots. K-means clustering found three categories that substantively described the data (see Figure IV.3). We found three of the four clusters described previously: LALE, HALE, and HAHE. Increasing the number of clusters did not result in more meaningful clusters but resulted only in additional dissections of Agency into three categories or more. 1316 (63.3%) ratings fell into the LALE category, 659 ratings (31.7%) fell into the HALE category, and 104 ratings (5%) fell into the HAHE category. Clustering on the principal components found highly similar results: 1307 (62.9%) ratings fell into the LALE category, 658 ratings (31.6%) fell into the HALE category, and 114 ratings (5.5%) fell into the LAHE category. No adult thought any robot was LAHE; if robots had high experience they necessarily also had high agency. All subsequent analyses will be based on the CFA clusters for brevity because analysis of the PCA clusters produced nearly identical results.

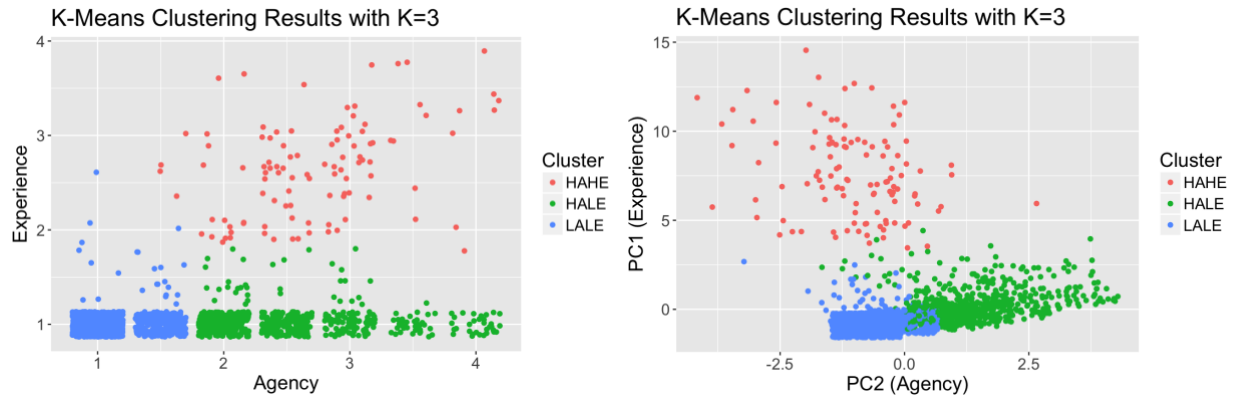


Figure IV.3. A Visualization of K-Means Clustering for Adult Responses. A visualization of how k-means clustering divided participants responses into three categories: high agency, high experience (HAHE); high agency, low experience (HALE); and low agency, high agency (LAHE). The left panel shows clustering performed on ratings of factors. The right panel shows clustering performed on principal components. Responses were jittered for visibility.

Characteristics of robots. We used the three identified categories of mental attributions to determine how robot characteristics impacted attributions of mental abilities to robots. We inspected whether specific robots uniformly fell into only one category of mental attributions or multiple. We found that responses for each robot did not fall uniformly nor exclusively into individual clusters. Each robot had responses that fell into each of the three clusters. In other words, the same robot was not viewed similarly by all participants: Different participants attributed different levels of mental abilities to the same robot.

To clarify and complement these robot ratings and groupings, we then examined the distribution of responses for each robot across the three types of mental attributions (see Table 1 and refer to Figure IV.1 for identifying pictures of each robot). Here we measured the proportion of responses that fell into each category for each robot. We then performed K-means clustering on the resulting distribution of proportions to see if we could uncover groups of robots that shared similar distributions of ratings. Three categories of robots were extracted. Sofia and

Actroid shared similar distributions of responses across the three types of mental attributions. Atlas and Spot were also clustered together. The remaining six robots were grouped together.

Table IV.1

<i>Proportion of Responses in Each Type of Mental Attribution</i>				
Robot	HAHE	HALE	LALÉ	Cluster
Festo	.0174	.300	.683	Mechanical
Kaspar Back	.0342	.205	.761	
Kaspar Front	.0714	.223	.705	
Nao	.0351	.268	.697	
Pepper	.0388	.284	.677	
Tapia	.0606	.212	.727	
Actroid	.0783	.326	.596	Human-Like
Sofia	.0733	.332	.595	
Atlas	.0556	.470	.474	Situationally-Aware
Spot	.0386	.442	.519	

Note: Distribution of responses for each robot into the three categories of mental attributions and their clusters (identified by K-means clustering).

Based on a qualitative viewing of the robots in each category, we labeled each cluster to more clearly describe and differentiate cluster membership. The *Mechanical* cluster contains robots with clearly visible mechanical components: wires, silicon or plastic exteriors, or non-human shapes. The *Human-Like* cluster contains the most closely human-like robots of the sample. Finally, the *Situationally-Aware* cluster contains the only two robots that performed behaviors consistent with being aware of objects in their environment: Atlas pursued a moving box and Spot avoided an obstacle by ducking under a table.

Individual differences and consistencies. We also determined whether the responses of a specific participant were consistent across different robots (i.e., their responses fell exclusively into only one category of mental attributions regardless of robot, e.g. LALÉ) or whether they varied across robots (i.e., their responses fell into different categories according to robot).

Ultimately, we found that participant responses were more likely to fall into a single category. For 254 participants (59.6%), *all* of their responses fell into only a single category. Of these participants whose responses fell into a single category, 185 participants' responses were LALE, 58 participants were HALE, and 11 responses were HAHE. For 158 participants (37.1%), their responses fell into 2 clusters. When participants' responses fell into two categories, 147 participants' (93.0%) responses were all HALE or LALE, 6 participants were LALE or HAHE, 5 participants were HALE or HAHE. Only 9 participants' responses (2.1%) fell into all 3 categories.

Predicting mental attributions. Participant and Robot Cluster were explored for their impact on attributions of mental abilities to robots. We used the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015) to perform a linear mixed effects analysis of the relationship. As fixed effects, we entered Robot Cluster, Age and Gender (without interaction terms) into the model. As random effects, we had intercepts for participant and robot.

When predicting Experience, subject-to-subject variability had the greatest contribution among the random effects. The estimated variance (SD = .336, 95% CI [.313, .362]) for subject-to-subject variability was nearly twice as much as the within-subject variability (SD = .178, 95% CI [.172, .185]). In other words, participants' responses were more consistent across robots than they were between participants. Moreover, robot-to-robot variability had the smallest contribution to Experience (SD = .000, CI 95% [.000, .017]). According to likelihood-ratio tests (LRT), Robot Cluster significantly predicted Experience: The *Human-Like* cluster had significantly higher ratings of Experience compared to the *Situationally-Aware* cluster, $\beta = .038$, CI 95% [.012, .064], and the *Mechanical* cluster, $\beta = .044$, CI 95% [.022, .066], $\Delta\chi^2(2) = 25.9$, $p < .001$. There was no difference in Experience between the *Situationally-Aware* and *Mechanical*

cluster, $\beta = -.005$, CI 95% [-.028, .016]. Fixed effects explained 3.3% of the variance, marginal $R^2 = .033$, compared to 78.7% explained by both fixed and random effects, conditional $R^2 = .787$ (Bartoń, 2018; Nakagawa & Schielzeth, 2013). Therefore, there are unmeasured variables related to participant that are likely better predictors of Experience than the fixed effects measures included in our model.

When predicting Agency, subject-to-subject variability had the greatest contribution (SD = .613, CI 95 % [.569, .661]). The within-subject variability (SD = .441, CI 95% [.427, .457]) and robot-to-robot variability (SD = .029, CI 95% [.000, .067]) contributed less. Just as with Experience, there was more variability between participants ratings of Agency than there were between robots. An LRT indicated that Robot Cluster significantly predicted Agency: The *Situationally-Aware* cluster was rated as having more Agency than both the *Human-Like* cluster, $\beta = .188$, CI 95% [.095, .281]. and the *Mechanical* cluster, $\beta = .391$, CI 95% [.315, .469], $\Delta\chi^2(2) = 25.9$, $p < .001$. The *Human-Like* cluster had significantly higher ratings of Agency than the *Mechanical* cluster, $\beta = .202$, CI 95% [.127, .280], $\Delta\chi^2(2) = 25.9$, $p < .001$. The variance explained by random effects was substantially larger than that measured by fixed effects, marginal $R^2 = .051$, conditional $R^2 = .676$ (Bartoń, 2018; Nakagawa & Schielzeth, 2013).

Robot uncanniness. We performed a linear mixed effects analysis to assess the impact of mental attributions, Participant and Robot Cluster on ratings of Uncanniness. As fixed effects, we entered Agency, Experience, Robot Cluster, Age, and Gender (without interaction terms) into the model. As random effects, we had intercepts for participant and robot.

When predicting Uncanniness, both subject-to-subject variability (SD = .542, CI 95% [.496, .592]) and between-robot variability (SD = .349, CI 95% [.234, .588]) contributed less than within-subject variability [SD = .654, 95% CI [.633, 677]]. There was more variation within

participants' Uncanniness responses than between participants or robots. LRTs indicated that increases in Experience, $\beta = .066$, CI 95% [.016, .115], $\Delta\chi^2(1) = 6.69$, $p = .009$, and Agency, $\beta = .105$, CI 95% [.059, .151], $\Delta\chi^2(1) = 20.2$, $p < .001$, predicted increases in Uncanniness. Also, the *Human-Like* cluster was considered more uncanny than the *Situationally-Aware*, $\beta = .766$, CI 95% [.003, 1.529], and *Mechanical* cluster, $\beta = .855$, CI 95% [.231, 1.477], $\Delta\chi^2(2) = 6.41$, $p = .041$. Nevertheless, caution is warranted when interpreting these fixed effects. When the sample size is large ($N = 2079$), the likelihood of finding a statistically significant difference increases although there may not be a practically significant difference. The estimates and confidence intervals for the coefficients for Agency and Experience were very close to 0. They, therefore, may not be strong predictors of Uncanniness regardless of statistical significance. Fixed effects explained 17% of the variance, marginal $R^2 = .170$. Fixed effects and random effects combined explained 57.9% of the variance, conditional $R^2 = .579$ (Bartoń, 2018; Nakagawa & Schielzeth, 2013).

Discussion

Mental attributions to robots. Adults model the minds of robots into two components of mental abilities: psychological agency and perceptual experience, as demonstrated by both a CFA and PCA. Moreover, adults attribute these abilities in distinctive ways. More often than not robots were considered to have low agency and low experience (63.3% of ratings fell in this category). When adults claimed that a robot had high agency, that robot could also either have high experience (5%) or low experience (31.7%). No adult thought any robot had low agency and high experience, if robots had high experience they necessarily also had high agency

Characteristics of robots. Specific robots do not fall into single categories of mental attributions. In fact, it appears that features of the robots are not substantially important for

determining the attributed mental abilities of the robot. Robots that were shown to be situationally aware were rated as having more agency than mechanical and human-like robots. Therefore, behavior may be an important factor in determining agency. Sensibly, human-like robots (that looked very closely human) were also rated with most capacity for experience, more than the situationally aware and more mechanical robots. But, to reiterate, these effects may not be practically significant (effects were close to zero).

Individual differences and consistencies. Even though, as just discussed, characteristics of the robots did matter, stable differences in individual participants' ratings of robots surfaced as the most important factor for determining how a robot's mental abilities were rated. Most adults' responses typically fell into a single category. If an adult reported that one robot had low agency and low experience (LALE), he or she was highly likely to report that the other robots in our sample of robots did as well. Moreover, our linear mixed effects model found that as-yet-unmeasured variables related to the participant were more likely to explain participants' mental attributions to robots than the robots themselves.

Robot uncanniness. In this sample, mental abilities did help account for a statistically significant portion of the variance in uncanniness, as it has in prior research contrasting two or three robots at a time. Moreover, the human-like robots were considered the most uncanny compared to the mechanical and situationally aware robots, consistent with past findings. In our study, however, caution must be taken when interpreting these effects. Although the effects were statistically significant, the effects of mental attributions may not strongly explain ratings of uncanniness (their estimates of coefficients were small and approaching zero). Further, including participants' stable tendencies to regard the robots as generally similar (e.g. all as LALE) greatly increased the total variance accounted for in participants' Uncanny judgments. In essence,

participants who saw the robots as LALE also rated them as less uncanny than participants who saw them as HAHE. Still, unmeasured variables related to the participant are likely also important for predicting uncanniness. Future studies should include more robots, as we have done, but also include ratings of additional factors, including factors that characterize the participants themselves, that may better explain variations in uncanniness. We speculate on some potentially informative additional factors in the General Discussion that could be included in future research, including testing a wider age span of participants.

Study 2: Children

Children are more likely to attribute mental abilities to robots (Brink et al., 2017; Kahn et al., 2012) than adults. Moreover, previous research suggests that children attribute uncanniness differently to robots than do adults (e.g., Brink et al., 2017). Utilizing the same design as in Study 1, we investigated whether robot appearance or individual differences was more important for mental attributions and uncanniness for children, and we explored if and how those attributions changed with age.

This project (Study 2) is still currently in progress as we are still collecting child data. Nevertheless, we have collected data from over 100 children.

Method

Participants. 120 children (54 females), 3.41 to 17.46 years old, were recruited from a local natural history museum between February 2017 and May 2018. Children were questioned in a semi-isolated, quiet space within the museum. Seven children were excluded due to low concentration on the task. Because data were collected in a public space, we did not collect information regarding children's race, ethnicity, or socioeconomic status. Written parental consent and verbal child assent were obtained first. Children received a small toy for

participating.

Task and design overview. Children rated three videos of real-life robots—randomly selected from a set of 11 (see Figure IV.1 and Figure IV.4) videos of real-life robots—as well as the same video of a person used in Study 1. For each robot, children answered 10 questions concerning their feelings about the robot and their beliefs about the mental abilities of that robot. The study typically lasted less than 15 minutes.

Videos. Children viewed the same 10 robots that adults viewed in Study 1 plus one additional robot (see Figure IV.1 for the first 10 robots; see Figure IV.4 for the additional robot). The additional commercially available robot was a robot specially designed for interactions with humans.



Figure IV.4. The Additional Robot ASIMO, Included in Study 2.

Robot Beliefs Interview (RBI). The same nine interview questions used in Study 1 (taken from Brink et al., 2017, Robot Beliefs Interview) were used as in Study 1, but presented in a face-to-face interview format rather than online. These questions were validated for use with children in Brink et al. (2017) and again include items relevant to the aggregate of feelings of *Uncanniness* and attributions of *Psychological Agency* and *Perceptual Experience*.

Uncanniness. For each robot video, children were assessed on their uncanny response via two questions gauging the extent to which they felt the robot was creepy or unsettling: (1) “Do you feel the robot is nice or creepy?” and (2) “Does the robot make you feel weird or happy?” Utilizing the methods established in Brink et al. (2017) the response format resulted in a 4-point scale for each question coded as 0 (Nice/Happy), 1 (Creepy/Weird-a little bit), 2 (Creepy/Weird-a medium amount), and 3 (Creepy/Weird-a lot).

Mental attributions. Children also answered seven questions assessing the robots’ mental capacities for *Psychological Agency* and *Perceptual Experience*. Agency was captured with three items: “does the robot” (1) “choose to move?”, (2) “think for itself?”, and (3) “know the difference between good and bad?” Experience was captured by three items: “would the robot” (1) “feel pain?”, (2) “feel scared?”, (3) “feel hungry?” An additional exploratory question asked, “Does this robot have feelings, like happy and sad?” Each question resulted in a 4-point scale coded as 0 (No), 1 (Yes-a little bit), 2 (Yes-a medium amount), and 3 (Yes-a lot). (See Appendix IV.B for complete child interview.)

Robot validity check. We again included an additional item with the RBI to check whether children were properly aware of the nature of the stimuli as either a human or robot. Immediately after watching each video, children were asked to report whether they believed the agent in the video was a robot or not. After responding, the experimenter told the child the actual identity of the agent in the video, e.g., “You’re right. That is a robot” or “Actually, that’s a robot.”

Results

Research questions. Analyses mirrored those of Study 1 addressing: (1) ***Mental attributions to robots:*** Do children attribute mental abilities to real-world robots similarly to the

way that adults do? (2) **Characteristics of robots:** Do the characteristics of robots, like appearance and behavior, predict the types of mental attributions that children make and the strength of those mental attributions? And if so, how? (3) **Individual differences and consistencies:** Do children change how they attribute mental attributes to robots depending on the type of robot, or do individuals consistently apply the same mental attributions to robots regardless of robot? (4) **Robot uncanniness:** How do mental attributions impact children's uncanniness responses to a wide variety of real-world robots?

Data analysis plan. The data analysis plan was identical to that of Study 1. To begin we analyzed all child data together. Given the range of children's ages from 3 ½ to 17 ½ years, we later conducted several analyses separating younger and older children.

Imputation. We applied missing value imputation to the data set to attempt to increase statistical power. We utilized nonparametric missing value imputation using Random Forest (Stekhoven, 2013). The out-of-bag (OOB) imputation error estimate for our data was 32.61%.

All analyses were performed both with a fully imputed data set and the original data set. However, only the results from the original data set are described here. We describe only the original data set, because as the OOB error estimate shows, there was not sufficient data to improve the accuracy of the data by imputation methods.

Robot validity check. Six robots were identified as robots by more than 97% of the participants. Eight of 10 robots were identified by robots by more than 85% of participants. Actroid and Sofia were identified as robots by only 67.5% and 63.6% of participants respectively. The still images in Figure IV.1 (of Study 1) show that Actroid and Sofia are highly human-like in outer appearance. At same time, however, more than 94% of participants

identified the video of the person as a person. With this measure then, all 10 robots were highly robot-like to children.

Mental attributions to robots. *Confirmatory Factor Analysis.* We performed a CFA to determine whether children attribute mental abilities to a broad range of robots similarly to the way they have done in previous studies (Brink et al., 2017). A diagram of the CFA model can be seen in Figure IV.2.

This initial CFA model provided the same factor structure identified using EFA by Brink and colleagues (2017) with ratings of only three robots. The current model approached a good fit of the data. The four commonly used model fit indices were within or approaching their respective cutoff ranges (RMSEA = .08 < 0.05; CFI = 0.945 > 0.95; TLI = 0.910 > 0.95; SRMR = .052 < .08) (Hu & Bentler, 1999). All factor loadings met or were near the threshold for good indicators, .7 (Kline, 2011). For Agency, its three indicators had factor loadings between .564 and .735. The three indicators for Experience were within .592 and .828. For Uncanniness, the two indicators had loadings of .770 (creepy) and .711 (weird). More data should be collected (as is being done) to support that this model is a good fit of the data, confirming Brink et al. 2017.

From the CFA analysis, and paralleling the analyses of adult data (in Study 1), we created three aggregates: two types of mental attribution aggregates—Psychological Agency and Perceptual Experience—and an Uncanniness aggregate. Each aggregate was calculated by averaging participants' responses for each aggregate's respective items (as shown in Figure IV.2).

Principal Component Analysis. The PCA included information from one additional item that was not included in the CFA (so that the CFA could confirm Brink et al., 2017): *feel*. Again,

because we wanted to later predict uncanniness feelings, all uncanniness items were excluded from the PCA analysis.

The two components extracted from the PCA explained 65.3% of the variance in our data. The first principal component (PC1) equally weighted all seven items: *choose to move*, *think*, *feel*, *scared*, *hungry*, *pain*, and *moral*. This first principal component explained 49.2% of the variance. The second principal component (PC2) placed more weight on the subset *pain*, *scared*, *moral*, *choose to move* and explained 16.1% of the variance. Thus, the PCA ultimately produced substantively and quantitatively distinct types of mental attributions compared to the CFA. The first component encompassed all interview items and thus may represent general mental ability as children judge it for robots. The second component might be interpreted as moral agency and moral concern.

The principal components calculated by the PCA were utilized in subsequent analyses but also compared to the results for the CFA.

Clustering. The two groups of aggregates, the factors and principal components resulting from the CFA and PCA, were analyzed using unsupervised statistical learning methods (i.e., K-means clustering and hierarchical clustering) just as done with the adult data in Study 1.

For our 113 participants, we obtained 288 ratings across 12 robots. K-means clustering found that four categories substantively described the data (see Figure IV.5): LALE (i.e. LoAgency-LoExperience), HALE, HAHE, and LAHE (the LoAgency-HiExperience cluster never utilized by adults). Matching the pattern found with adults, 161 ratings (55.9%) fell into the LALE category, 67 ratings (23.3%) fell into the HALE category, 33 ratings (11.5%) fell into the HAHE category. Additionally, 27 ratings (9.4%) fell into the LAHE category. Whereas all

adults thought that if a robot had high experience it necessarily also had high agency, this was not true for some children.



Figure IV.5. A Visualization of K-Means Clustering for Child Responses. A visualization of k-means clustering for child responses into four categories: low agency, low experience (LALE); high agency, low experience (HALE); high agency, high experience (HAHE); and low agency, high experience (LAHE).

Characteristics of Robots. We then used these categories of mental attributions to determine how robot characteristics impacted attributions of mental abilities to robots. Thus, we inspected whether specific robots typically fell into only one category of mental attributions or multiple. Just as for adults, we found that responses for each robot did not fall uniformly and exclusively into a single category. Altogether participants attributed different mental abilities to the same robot.

We then examined the distribution of responses for each robot across the four types of mental attributions. We measured the proportion of responses that fell into each category for each robot. We then performed K-means clustering and hierarchical clustering, attempting to

define 2 to 5 clusters, on the resulting distribution of proportions. Unfortunately, the robots did not divide into substantively informative clusters. Hopefully, with more data, clear categories of robots will appear.

Individual differences and consistencies. We also determined whether the responses of a specific participant were consistent across different robots or whether they varied across robots. Children's responses were most likely to fall into a single cluster. For 64 participants (59.3%), all of their responses fell into only a single cluster. For 41 participants (38.0%), their responses fell into 2 clusters. Only 3 participants' responses (2.8%) fell into all 3 categories and no participants' responses fell into all 4 categories. In these ways, children's responses closely fit the exact same patterns as adults.

Of the participants whose responses fell into only a single category, 45 of children's responses were LALE, 12 were HALE, 5 responses were HAHE, and 2 were LAHE. Similarly, for adults, recall that for participants whose responses fell into a single category, the top two frequencies were clearly LALE followed by HALE.

Predicting mental attributions. Participant and Robot were explored for its impact on attributions of mental abilities to robots. We used the lme4 package in R (Bates et al., 2015) to perform a linear mixed effects analysis of the relationship. As fixed effects, we entered Age and Gender (without interaction terms) into the model. As random effects, we had intercepts for participant and robot.

When predicting Experience, subject-to-subject variability had the greatest contribution among the random effects. The estimated variance ($SD = .604$, 95% CI [.508, .717]) for subject-to-subject variability was greater than the within-subject variability ($SD = .485$, 95% CI [.434, .545]). and robot-to-robot variability ($SD = .059$, CI 95% [.000, .200]). So, as for adults, child

participants' responses were more consistent across robots than they were between participants. According to LRTs, increases in Age significantly predicted decreases in ratings of Experience, $\beta = -128$, CI 95% [-.173, -.083], $\Delta\chi^2(1) = 27.8$, $p < .001$. This is consistent with the findings from Brink et al. 2017 (see Chapter II) for child rating of three robots. Gender did not significantly predict Experience. $\Delta\chi^2(1) = .3$, $p = .58$. Fixed effects explained 19.3% of the variance, marginal $R^2 = .193$, compared to 68.6% explained by both fixed and random effects, conditional $R^2 = .686$ (Bartoń, 2018; Nakagawa & Schielzeth, 2013). Age is, therefore, an important factor in predicting ratings of Experience. Furthermore, there are other, as-yet-unmeasured variables related to participant that could improve prediction of ratings of Experience.

When predicting Agency, subject-to-subject variability had the greatest contribution (SD = .684, CI 95 % [.569, .817]). The within-subject variability (SD = .575, CI 95% [.514, .649] and robot-to-robot variability (SD = .164, CI 95% [.000, .356]) contributed less. As for Experience and as for adults in Study 1, there was more variability between participants' ratings of agency than there were between robots. According to LRTs, increases in Age significantly predicted decreases in ratings of Agency, $\beta = -136$, CI 95% [-.188, -.085], $\Delta\chi^2(1) = 24.4$, $p < .001$. This too is consistent with the findings from Brink et al. 2017 (see Chapter II) for child ratings of three robots. Gender did not significantly predict Agency, $\Delta\chi^2(1) = .27$, $p = .61$. Fixed effects explained 16.6% of the variance, marginal $R^2 = .166$, compared to 66.6% explained by both fixed and random effects, conditional $R^2 = .666$ (Bartoń, 2018; Nakagawa & Schielzeth, 2013). Age is, therefore, an important factor in predicting ratings of Agency. There are likely additional unmeasured variables related to participant that could improve prediction of ratings of Agency.

Robot uncanniness. We performed a linear mixed effects analysis to assess the impact of mental attributions, Participant and Robot Cluster on ratings of Uncanniness. As fixed effects,

we entered Agency, Experience, Robot Cluster, Age, and Gender (without interaction terms) into the model. As random effects, we had intercepts for participant and robot.

When predicting Uncanniness, both subject-to-subject variability (SD = .340, CI 95% [.000, .512]) and robot-to-robot variability (SD = .448, CI 95% [.274, .789]) contributed less than within-subject variability [SD = .796, 95% CI [.711, .900]]. LRTs did not find a significant relation between Experience or Agency and Uncanniness. Agency's impact on Uncanniness, however, was approaching significance, $\beta = -.123$, CI 95% [-.257, .008], $\Delta\chi^2(1) = 3.4$, $p = .065$. Fixed effects explained 1.5 % of the variance, marginal $R^2 = .015$. Fixed effects and random effects combined explained 34.3% of the variance, conditional $R^2 = .343$ (Bartoń, 2018; Nakagawa & Schielzeth, 2013). More data should be collected (as is being done) to strengthen this analysis.

Comparing younger and older children. In this sample, there were 77 children under the age of 9 and 44 children that were 9 and older, 9 years being the age where uncanniness feelings significantly shifted in Brink et al. (2017; see Chapter II). Children younger than 9 reported that robots with more mind were *less* creepy, whereas children older than 9 and adults reported that robots with more mind were *more* creepy. The data for younger children strongly resembles the data presented here in Study 2. The data for older children, however, began to approximate the adult data from Study 1. For example, the CFA model for older children appeared to be a stronger fit when the moral item was removed. There are not, as yet, sufficient data, however, to support this claim statistically. The PCA also began to resemble that of adults. The first component more heavily weighted *feel*, *hungry*, *pain*, and *scared*, whereas the second component more heavily weighted items linked *think* and *choose*. Moreover, older children's

responses appeared to fit three categories of mind—LALE, HALE, and HAHE—rather than four.

With fewer participants (n=44, 9 years and older) and each child rating only three robots, these findings are necessarily preliminary. More data are clearly needed (and still being collected) to confirm the differences and similarities between younger children, older children, and adults.

Discussion

Mental attributions to robots. In general, children model the minds of robots into two components of mental abilities: psychological agency and perceptual experience, as demonstrated by CFA. More often than not children considered robots to have low agency and low experience (63.3% of ratings fell in this category). When adults claimed that a robot had high agency, that robot could also either have high experience (5%) or low experience (31.7%). Some children, unlike adults, however, also reported that robots could fall into low agency and high experience. No adult thought any robot had low agency and high experience.

Characteristics of robots. Like for adults, specific robots did not fall into single categories. The features of robots were not the most important criteria for determining how to attribute mental abilities to robots. Instead, child differences were more critical for determining whether a robot has mental abilities.

Individual differences and consistencies. Most children's responses fell uniformly and consistently into a single category: If a child reported that one robot had low agency and low experience, they typically reported that other robots regardless of appearance also had low agency and low experience. We also found that one child-related measure in our data predicted

ratings of agency and experience: the age of the child. As age increased, ratings of agency and experience for robots decreased.

Robot uncanniness. In this sample, mental abilities did not explain most of the variance in uncanniness. Increases in agency were trending toward predicting decreases in uncanniness, similar to previous findings by Brink and colleagues (2017). Regardless, when children are given a larger range of robots to consider, something more complicated is going on than human-likeness or mental attributions than discussed in previous literature on robot uncanniness (K. Gray & Wegner, 2012).

Changes according to age. Preliminary data suggest that children's responses to these robots do in fact differ according to age. Younger children (younger than the age of 9) follow the pattern outlined above while older children (9 and older) begin to more closely resemble adults. However, more data are still needed to confirm these differences.

General Discussion

Across the two studies, we found that adults and children attribute mind to robots but in slightly different ways. Moreover, characteristics unique to the child or adult are important for how they attribute minds to robots. For adults, whereas there is some evidence that the behavior and appearance of the robot contribute to mental attributions, those do not predict mental attributions as well as the person themselves. This is true for not only mental attributions but also feelings of uncanniness. Due to limited data on children's responses, however, it is difficult to characterize the relations between child-specific variables, mental attributions, and feelings of uncanniness currently. Nonetheless, these findings, especially those for adults for now, have important implications for research on anthropomorphism, human-robot interaction, and robot design.

Adults divide mental abilities into two broad categories: Psychological Agency and Perceptual Experience. Robot behaviors, like goal-directed actions or situational awareness, can contribute to perceptions of psychological agency in a robot. Human-likeness in a robot may also contribute to attributions of perceptual experience. These characteristics of a robot, however, do not appear to have as strong an impact on mental attributions as do characteristics unique to the person viewing the robot. Those person-specific characteristics are as of yet unidentified. Future research should take more stock of individual difference measures in order to identify the mechanisms that produce mental attributions in nonhuman agents.

Children also divide the mind into Psychological Agency and Perceptual Experience. Children's versions of these constructs, however, may vary slightly from the adult versions. Children seemingly place value in moral agency and concern when characterizing agency. At the same time, like with adults, child-specific variables are likely the more important factors determining mental attributions and feelings of uncanniness toward robots. Mostly clearly, for now, children's age was predictive of mental attributions to robots: as age increased, children's attributions of agency and experience to robots decreased. More data need to be collected, however, to view more fine-detailed age differences and relations between robot characteristics, individual differences, mental attributions and feelings of uncanniness.

These results have important implications for understanding the mechanisms that impact anthropomorphism. Epley and colleagues (2007) theorized that person-specific characteristics are important in determining whether a person will attribute human-like qualities to a nonhuman agent. Our study supports this assertion. Individual differences in adults and children explained a substantial portion of variance in mental attributions to robots. Our research, however, did not explore the person-specific variables that contribute to these differences. Future studies should

more clearly investigate what the person brings to human-robot interactions. Clearly, the willingness of an adult or child to attribute mental states to a robotic agent is influenced by their own individual experiences.

Unfortunately, due to the anonymous nature of our surveys with adults, very few participant-specific variables were collected. Therefore, we were unable to explore the individual difference factors that lead to differences in attributions of mental abilities to robots. We found evidence of age-related differences in children but not the mechanism by which age predicts differences in uncanniness and mental attributions for robots. Important differences to consider may include the child's cohort, their experience with technology, cultural opinions, and expectations about technology and social behaviors. Future studies should explore these characteristics and how they impact mental attributions and uncanniness for robots.

Such research will be crucial for designers and roboticists focused on human-robot interaction. Due to the importance of person-specific variables in mental attributions and feelings of uncanniness for robots, designers and engineers should be very clear on the population for whom they are designing robots. Not only is the design of the robot, its behaviors and appearance, important, but also the person viewing the robot.

Given the current findings, clearly individual differences are particularly important in determining reactions to distinct robots and should be explored more in the future. We need more focus on individual characteristics of the child and adult interacting with the devices. What are the person-specific characteristics that impact mental attributions to and feeling about robots? What are the person-specific mechanisms that shape those attributions and feelings?

As one example, assume that child and adult experience with technology, or even specifically with robots, proves to be a key factor influencing their attributions to and feelings

about robots (which a priori seems likely). The current data suggest that these kinds of participant-specific factors are powerful enough that robots should come with pre-packaged experiential exercises to help their users approach the robots as the designers hope they will. The designer's own intuitions about what a robot evokes may very well be a poor source of information on this score, considering that robot designers by training have many, many hours of accumulated experiences and hopes about the robot they create. Empirical evidence would be a better and important source of information in this regard, especially for children.

Appendix IV.A

Robot Beliefs Interview - Adults

1. Is this a robot or a person?

Possible Responses to the following questions: “No,” “Yes – A little bit,” “Yes – A medium amount,” “Yes – a lot”)

2. Do you feel this robot is creepy?
3. Does this robot make you feel weird?
4. Does this robot purposefully choose to move?
5. Does this robot think for itself?
6. Does this robot know the difference between right and wrong?
7. If someone pinched the robot, would it feel pain?
8. If the robot saw a snake, would it feel scared?
9. If the robot did not eat breakfast, would it feel hungry?
10. Does the robot have feelings, like happy and sad?

Appendix IV.B

Robot Beliefs Interview - Children

1. Is this a robot or a person?
2. Do you feel the robot is nice or creepy?
 - 2.1. “Do you feel the robot is nice (thumbs up) or creepy (thumbs down)?”
 - 2.2. “How creepy do you feel it is? A little bit, a medium amount, or a lot?”
3. Does the robot make you feel weird or happy?
 - 3.1. “Does the robot make you feel weird (thumbs down) or happy (thumbs up)?”
 - 3.2. “How weird does it make you feel? A little bit, a medium amount, or a lot?”
4. When the robot moves, does it choose to move?
 - 4.1. “When the robot moves, does it choose to move?”
 - 4.2. “How many things can the robot choose to do? A few things, a medium amount of things, or a lot of things?”
5. Does the robot think for itself?
 - 5.1. “Does the robot think for itself?”
 - 5.2. “How much does it think for itself? A little bit, a medium amount, or a lot?”
6. Some actions are bad, like hitting. And some actions are good, like helping. Does this robot know the difference between good and bad?
 - 6.1. “Does this robot know the difference between good and bad?”
 - 6.2. “How much does it know the difference between good and bad? A little bit, a medium amount, or a lot?”

7. If I pinched the robot, would it feel pain?
 - 7.1. "If I pinched the robot, would it feel pain?"
 - 7.2. "How much can this robot feel pain? A little bit, a medium amount, or a lot?"
8. Does the robot have feelings, like happy and sad?
 - 8.1. "Does the robot have feelings, like happy and sad?"
 - 8.2. "How much does the robot have feelings? A little bit, a medium amount, or a lot?"
9. If the robot saw a snake, would it feel scared?
 - 9.1. "If the robot saw a snake, would it feel scared?"
 - 9.2. "How much can the robot feel scared? A little bit, a medium amount, or a lot?"
10. If the robot did not eat breakfast, would it feel hungry?
 - 10.1. "If the robot did not eat breakfast, would it feel hungry?"
 - 10.2. "How much can the robot feel hungry? A little bit, a medium amount, or a lot?"

Chapter V

As robots and technological devices increasingly appear to have human-like minds, it presents the opportunity to address several intriguing questions: (1) To what extent do we perceive that these devices share similarities with people and to what extent do we believe they have human-like minds? (2) Does an understanding of human-like qualities in a device affect our ability to understand and interact with it? (3) Does this understanding influence our willingness to accept information from these devices? And, of crucial importance, (4) how do answers to these questions vary with age?

In Chapter II, I reported research where children viewed two distinct robots, one very human-like and one machine-like, and reported on their feelings for these robots and their beliefs about the human-like minds of these devices. We found that as younger children attributed more mind to these robots, they experienced fewer feelings of uncanniness toward those robots. However, for older children, their feelings and beliefs about robots differed and their perceptions of the mental capacities of robots differentially impacted their feelings toward them—older children did not prefer robots that appeared to have mental abilities, similarly to adults who find machines with closely human-like minds to be unsettling.

In Chapter III, I presented research where 3-year-old children learned the names of novel objects from either a pair of social robots or a pair of inanimate machines. One informant in each pair was previously shown to be accurate and the other inaccurate. In these studies, children trusted information from an accurate social robot over an inaccurate one and even more so when they perceived the robots as having psychological agency. Children did not trust information

from the inanimate, but accurate, machine however. Children can learn from technological devices, e.g., social robots, but trust more when those devices appear to them to have mindful agency.

In Chapter 4, I described research where adults and children reported on their feelings for robots that differed on a wide variety of features. We found that on top of the previously identified variables important in predicting ratings of uncanniness—robot type, experience and agency—person-specific variables are likely also critical factors. For example, in our child study, age was an important factor in predicting mental attributions to robots. More research is needed to identify how individual differences explain both feelings of uncanniness and also mental attributions to robots.

Altogether, these results indicate that adults and children alike attribute qualities of a human-like mind to these devices. However, children and adults attribute mind to different extents and, when they do, it leads to different reactions to robots. Young children are more likely to attribute more human-like abilities to robots, including a larger range of robots (Chapter 4) and when they do, they like these devices more and trust them more easily. The willingness to attribute a human-like mind to a robot decreases, however, with age. Older children and adults are less likely to attribute human-like abilities, like perceptual experience, to robots. Moreover, adults and older children lose their affinity for these devices when attributions of mind increase (the opposite of young children's reactions). Finally, these changes may be more closely linked to characteristics of the child or adult, themselves, than the actual features of the robot. More research is needed to determine the exact individual characteristics of the person viewing the robot that impact beliefs and feelings about robots.

At the same time, child age seems clearly important. Although it is unclear what sorts of experiences and developments age is serving as a proxy for in robot research, our research nonetheless demonstrates that not only do children's beliefs about technology change with age but also that these differences in understanding directly impact children's interactions with this technology. For young children, when they perceived robots to have psychological agency, they used the robots' accuracy to determine which robot to trust for new information about novel objects. Young children expect and even prefer robot instructors that appear to have minds of their own, whereas older children do not expect robots to have minds and, I hypothesize on the basis of unsystematic research from others, are more indifferent to interactive and agent-like robot instructors.

Related Research on Robot Instructors

The set of child studies in this dissertation demonstrates the importance of psychological agency in determining whether young children trust social robots. However, this research as well as the broader literature demonstrate that older children think and feel differently about the psychological agency of robots. Whereas younger children in our studies were more likely to prefer and learn from robots with psychological agency, older children did not prefer these robots. Therefore, it is reasonable to expect that older children learn differently from instructional robots with agency compared to younger children. Indeed, existing research on the effectiveness of robot teachers for older children suggests that robots with more signals of agency, although effective educators for young children, show little or no improvements in learning for older children.

Several studies demonstrate that signs of agency in robots are less effective for encouraging learning in older age groups compared to younger age groups. In a table-setting

task, younger children learned better from a robot with a more socially interactive and communicative style of teaching than older children (Okita et al., 2009). Children 4- to 6-year-old improved so dramatically when taught by a robot with an interactive teaching style that they performed on par with 7- to 10-year-olds. There was no improvement in learning, however, for interactive teaching styles for the 7- to 10-year-olds compared to other less agent-like teaching styles. Further, while a human-like voice was better for accuracy in general, the effect was more pronounced for younger children. Older children did not show as much improvement as younger children when learning from a robot with a human voice compared to a monotone voice. In another study, the robot Robovie travelled around a Japanese school to speak English with 1st graders (6- to 7-year-olds) and 6th graders (11- to 12-year-olds) (Kanda, Hirano, Eaton, & Ishiguro, 2004). This robot performed interactive behaviors like hugging, shaking hands, playing rock-paper-scissors, singing, briefly conversing, and pointing at nearby objects. First graders spent significantly more time interacting with the robot than sixth graders did. Features of agency appear to be a more important factor for younger children when learning from robots.

These studies with older children, while informative, have not directly measured the impact of attributions of agency on their learning. More research is needed to evaluate how robots' features and children's developing cognitive abilities interact to produce improved learning in children from robots. We have begun to conduct such research. Some of that research evaluates the relationship between certain features of robots, and children's perceived overall sense of the robot as well as their sense of the agency of the robot. Ultimately, beyond these initial studies, future studies should investigate how children's understanding of robots impacts their perceptions of agency, the "worthiness" of robots as sources of information, and

consequently the quality of children's interactions with and learning from robots as children grow older.

Future Directions

In total, our research as well as others' has only just begun to demonstrate how children's beliefs and expectations about robots, their developmental trajectory with respect to this understanding, and the robot's design together impact how children will feel toward and learn from robots. Future research should continue to explore each of these components of child-robot interactions with more precision and within different contexts.

For instance, research should more fully identify the fine-grained developmental changes that occur in children's conceptual understanding of, and interactions with, robots. The research thus far, our own and that of others, only captures the emergence of these developmental changes on a very crude timeline, before and after approximately 9-12 years of age. One thing needed is more extended data (as in our Uncanny study) of a range of ages that help us and the field to move beyond the description of "younger" and "older" age groups to a more fine-grained set of developmental findings and hypotheses.

In line with this more comprehensive approach for exploring developmental changes in children's interactions with robots, future research should also investigate and clearly define the *mechanisms* that produce the child-adult differences that we identify here. The child-adult differences we have uncovered may very well be the result of developments in basic aspects of cognition as we have discussed above. Or they could result from younger children's increased interactions with more sophisticated and social technology on a regular basis compared to older children and adults (Bernstein & Crowley, 2008). That is, today's young children may grow up to look quite different from current adults because of different experiences with increasingly

different technological devices, including robots themselves, from adults and older children. It remains unclear how children's cognitive development in addition to their experiences with robots and smart technologies may interact to influence their interactions with these devices. It would be well worth following children (for example those in our sample) longitudinally to examine such intriguing possibilities.

As we continue to more precisely explore children's developmental trajectories and the mechanisms that determine how children will interact with technology, we can and should also consider child-robot interactions in special populations. Not only are typically developing children encountering new and different technology daily, but so too are children with special needs. Children in these populations bring their own set of unique experiences and cognitive abilities to their interactions with robots that may produce different outcomes from typically developing children. Indeed, research with children with ASD shows that they prefer robots that are the least human-like (Robins et al., 2006) unlike the children in our research who preferred the humanoid Nao robot over the mostly machine-like robot shown in Figure II.2. Given that robots are participating in educational endeavors with children with special needs, future studies should investigate the impact of child-specific abilities and experiences on child-robot interactions in these special populations.

The importance of child-specific features in child-robot interactions also highlights the informative potential of collaborative investigations between social scientists, engineers, and designers. Robot design currently, even those designed for use with children, has primarily been in the hands of engineers and corporate design teams. These designs are inspired, and limited, by the intuitions of these individuals and what they assume would be child-friendly and effective for children. This is similarly true for those robots increasingly found in movies for children. These

groups have their own marketing and research teams and as a result there is currently little collaboration between developmental researchers, engineers, designers, and movie animators. Collaboration would allow for developers and animators to share audience reaction information and product development testing which have been used to shape and develop their ideas and designs. Additionally, researchers could aid designers by sharing work on the interactions between children's developmental trajectories and robot design that would impact children's social and educational outcomes. For example, from our research we would likely conclude that designers should consider the effect of utilizing a variety of agent-like cues (e.g., contingent behaviors or abstractly human-like features) when designing educational robots. A collaboration among these groups could drastically expand our understanding of children's conceptual and cognitive development and simultaneously improve robot design for children by implementing more evidence-based decisions.

The rising number of robots interacting with children also has important implications for a variety of other developmental outcomes that I have not yet addressed across these research studies. Thus far, I have focused primarily on a small set of outcomes—how children feel toward robots and how children learn from robots—but social robots also have the potential to impact both children's social and moral development.

There is growing concern in the field that increased interactions with smart technologies, including social robots, might adversely impact children's social and moral development. Smart technology and social robots are frequently blamed for decreasing in-person social interactions, preventing children from learning how to effectively interact with others, and thus hindering children's social development (Turkle, 2011). Researchers, parents, and teachers are particularly concerned that interactions with robots will promote the development of antisocial behaviors. In

fact, there is some evidence for this. A hitchhiking robot that had successfully traveled around Germany, Canada and the Netherlands taking pictures and carrying on conversations with other travelers was eventually vandalized and destroyed a few weeks into its U.S. journey (Victor, 2015). A mall security robot designed to share information with customers routinely faced abuse from unsupervised children as they would often kick and push the robot (Brscić, Kidokoro, Suehiro, & Kanda, 2015; Nomura, Kanda, Kidokoro, Suehiro, & Yamada, 2017).

Nevertheless, empirical research suggests that these antisocial behaviors toward robots can be reduced and, moreover, that social robots can even be used to promote positive social development for certain populations. Researchers have found that, while abusive behavior toward robots does occur, a few behavioral modifications to robots can reduce this behavior. Preschool children in a classroom comforted a robot with a hug and protected it from other aggressive children when it started to cry after being damaged or played with too roughly (Carey & Markoff, 2010). Other studies show that children claim that a robot deserves to be treated fairly and not psychologically harmed after conversing and playing with the robot for 15 minutes (Kahn et al., 2012). Social robots have also been used to aid children with ASD in social development by practicing social behaviors like conversation with them and demonstrating common social cues and behaviors (Ricks & Colton, 2010). Future studies should continue to investigate the complex relationship between children's perceptions of robots, how they treat them, and ultimately how these interactions impact children's later social and moral development and their interactions with others.

Children increasingly encounter robots designed to comfort, teach, and play with them. It is therefore imperative to learn which robot features and child-robot interactions impact these interactions and outcomes. And, importantly, it is imperative to identify how and when these

interactions and outcomes change throughout childhood. Clearly, children's understanding of the mental capacities of robots appears as an important factor in their feelings toward and willingness to learn from robots. For these and many other reasons, more developmental research is badly needed to assess how children's cognitive abilities, their developmental trajectories, and the design of robots all work together to impact children's learning and feelings towards robots. Social robots are only one form of interactional, educational smart technology that may (or may not) promote these desired outcomes, but their special features and their increasing presence in children's lives make them worthy of study in their own right. Moreover, they allow for larger lessons on children's increasingly voluminous interactions with smart technological devices more generally.

Every year robots increasingly become a part of our lives and the lives of children. Investigations into children's understanding of and interactions with these devices is therefore only becoming more important. The future of research into child-robot interactions should continue to fully explore what children think and feel about robots, how they interact with them in a variety of contexts, and also expand those investigations to explore a wider variety of developmental outcomes for children, including educational, social, and moral outcomes.

References

- Bartoń, K. (2018). MuMIn: Multi-Model Inference. *R package version 1.40.4*. <https://CRAN.R-project.org/package=MuMIn>.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York, NY: Oxford University Press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01
- Beran, T. N., Ramirez-Serrano, A., Vanderkooi, O. G., & Kuhn, S. (2013). Reducing children's pain and distress towards flu vaccinations: A novel and effective application of humanoid robotics. *Vaccine*, *31*(25), 2772-2777. doi:http://dx.doi.org/10.1016/j.vaccine.2013.03.056
- Bernstein, D., & Crowley, K. (2008). Searching for signs of intelligent life: An investigation of young children's beliefs about robot intelligence. *The Journal of the Learning Sciences*, *17*(2), 225-247. doi:http://dx.doi.org/10.1080/10508400801986116
- Breazeal, C., Harris, P. L., DeSteno, D., Westlund, K., Jacqueline, M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in cognitive science*. doi:http://dx.doi.org/10.1111/tops.12192
- Brink, K. A., Gray, K., & Wellman, H. M. (2017). Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child development*. doi:http://dx.doi.org/10.1111/cdev.12999
- Brink, K. A., & Wellman, H. M. (submitted). Robot Teachers for Children? Young Children Trust Robots Depending on their Perceived Accuracy and Agency.
- Brscić, D., Kidokoro, H., Suehiro, Y., & Kanda, T. (2015). *Escaping from children's abuse of social robots*. Paper presented at the Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction.

- Carey, B., & Markoff, J. (2010). Students, meet your new teacher, Mr. Robot. *The New York Times*. Retrieved from <http://nyti.ms/1H9yiEN>
- Cohen, J. (1988). *Statistical power analysis for the social sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, 13(4), 148-153.
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 1149-1157.
- Dalecki, M., & Willits, F. K. (1991). Examining change using regression analysis: Three approaches compared. *Sociological Spectrum*, 11(2), 127-145.
doi:<http://dx.doi.org/10.1080/02732173.1991.9981960>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4), 864.
doi:<http://dx.doi.org/10.1037/0033-295X.114.4.864>
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.
- Glaser, A. (2016). Jibo is like Alexa and a puppy inside one adorable robot. Retrieved from <http://www.wired.com/2016/06/jibo-like-alexa-puppy-inside-one-adorable-robot/>
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619-619. doi:<http://dx.doi.org/10.1126/science.1134475>
- Gray, K., Knickman, T. A., & Wegner, D. M. (2011). More dead than dead: Perceptions of persons in the persistent vegetative state. *Cognition*, 121(2), 275-280.
doi:<http://dx.doi.org/10.1016/j.cognition.2011.06.014>
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125-130.
doi:<http://dx.doi.org/10.1016/j.cognition.2012.06.007>

- Hall, D., & Williams, C. (Writers). (2014). *Big Hero 6* [Motion Picture]. In. United States: Walt Disney Studios Motion Pictures.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental science*, *13*(6), 923-929.
- Han, J. (2010). *Robot-aided learning and r-learning services*: INTECH Open Access Publisher.
- Hanson, D. (2006). *Exploring the aesthetic range for humanoid robots*. Paper presented at the Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science.
- Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., . . . Stephanou, H. (2005). *Upending the uncanny valley*. Paper presented at the Proceedings of the national conference on artificial intelligence.
- Hashimoto, T., Kobayashi, H., Polishuk, A., & Verner, I. (2013). *Elementary science lesson delivered by robot*. Paper presented at the Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction, Tokyo, Japan.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243-259. doi:10.2307/1416950
- Horst, J. S., & Hout, M. C. (2014). *The Novel Object and Unusual Name (NOUN) Database: a collection of novel images for use in experimental research*. Unpublished Manuscript.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55. doi:http://dx.doi.org/10.1080/10705519909540118
- Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child development*, *78*(6), 1675-1688. doi:http://dx.doi.org/10.1111/j.1467-8624.2007.01095.x

- Jipson, J. L., Gülgöz, S., & Gelman, S. A. (2016). Parent–child conversations regarding the ontological status of a robotic dog. *Cognitive development, 39*, 21-35. doi:<http://dx.doi.org/10.1016/j.cogdev.2016.03.001>
- Johnson, C. N., & Wellman, H. M. (1982). Children's developing conceptions of the mind and brain. *Child development, 53*, 222-234. doi:<http://dx.doi.org/10.2307/1129656>
- Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in cognitive sciences, 4*(1), 22-28.
- Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., . . . Shen, S. (2012). “Robovie, you'll have to go into the closet now”: Children's social and moral relationships with a humanoid robot. *Developmental psychology, 48*(2), 303-314. doi:<http://dx.doi.org/10.1037/a0027033>
- Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-computer interaction, 19*(1), 61-84.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological science, 15*(10), 694-698. doi:10.1111/j.0956-7976.2004.00742.x
- Koenig, M. A., & Harris, P. L. (2007). The basis of epistemic trust: Reliable testimony or reliable sources? *Episteme, 4*(03), 264-284.
- Langlois, J. H., Roggman, L. A., & Rieser-Danner, L. A. (1990). Infants' differential social responses to attractive and unattractive faces. *Developmental psychology, 26*(1), 153. doi:<http://dx.doi.org/10.1037/0012-1649.26.1.153>
- Lasseter, J. (Writer). (1995). Toy Story [Motion Picture]. In. United States: Buena Vista Pictures.
- Low, A. (2016). Asus' adorable Zenbo home robot will be released in 9 to 12 months. Retrieved from <http://www.cnet.com/news/asus-zenbo-will-only-be-released-in-9-12-months/>

- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological science*, *16*(8), 601-608.
- MacDorman, K. F. (2006). *Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley*. Paper presented at the ICCS/CogSci-2006 Long Symposium: Toward social mechanisms of android science, Vancouver, Canada.
- MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, *25*(3), 695-710. doi:<http://dx.doi.org/10.1016/j.chb.2008.12.026>
- MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, *7*(3), 297-337. doi:<https://doi.org/10.1075/is.7.3.03mac>
- Mejías, C. S., Echevarría, C., Nuñez, P., Manso, L., Bustos, P., Leal, S., & Parra, C. (2013). Ursus: A Robotic Assistant for Training of Children with Motor Impairments. In J. L. Pons, D. Torricelli, & M. Pajaro (Eds.), *Converging Clinical and Engineering Research on Neurorehabilitation* (pp. 249-253). Berlin, Heidelberg: Springer.
- Melson, G. F., Kahn, P. H., Beck, A., & Friedman, B. (2009). Robotic pets in human lives: implications for the human–animal bond and for human relationships with personified technologies. *Journal of Social Issues*, *65*(3), 545-567.
- Michotte, A. (1963/2017). *The perception of causality* (Vol. 21): Routledge.
- Mikropoulos, T., Misailidi, P., & Bonoti, F. (2003). Attributing human properties to computer artifacts: Developmental changes in children’s understanding of the animate–inanimate distinction. *Psychology: The Journal of the Hellenic Psychological Society*, *10*, 53-64.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child development*, *78*(2), 622-646. doi:<http://dx.doi.org/10.1111/j.1467-8624.2007.01018.x>

- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98-100. doi:<http://dx.doi.org/10.1109/MRA.2012.2192811>
- Movellan, J., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). *Sociable robot improves toddler vocabulary skills*. Paper presented at the Proceedings of the 4th ACM/IEEE international conference on Human robot interaction.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.
- National Robotics Initiative 2.0: Ubiquitous Collaborative Robots (NRI-2.0). (2017). Retrieved from <https://www.nsf.gov/pubs/2017/nsf17518/nsf17518.htm>
- Nomura, T., Kanda, T., Kidokoro, H., Suehiro, Y., & Yamada, S. (2017). Why do children abuse robots? *Interaction Studies*, 17(3), 347-369.
- Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. (2009). *Learning together: ASIMO developing an interactive learning partnership with children*. Paper presented at the Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on.
- Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. K. (2011). Multimodal approach to affective human-robot interaction design with children. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1(1), 5.
- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental psychology*, 43(5), 1216.
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding statistics: Statistical issues in psychology, education, and the social sciences*, 2(1), 13-43. doi:http://dx.doi.org/10.1207/S15328031US0201_02

- Richert, R. A., & Harris, P. L. (2006). The ghost in my body: Children's developing concept of the soul. *Journal of Cognition and Culture*, 6(3-4), 409-427. doi:<http://dx.doi.org/10.1163/156853706778554913>
- Ricks, D. J., & Colton, M. B. (2010). *Trends and considerations in robot-assisted autism therapy*. Paper presented at the 2010 IEEE International Conference on Robotics and Automation (ICRA) Anchorage, AK.
- Robins, B., Dautenhahn, K., & Dubowski, J. (2006). Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies*, 7(3), 509-542.
- Schlottmann, A., Ray, E. D., & Surian, L. (2012). Emerging perception of causality in action-and-reaction sequences from 4 to 6 months of age: Is it domain-specific? *Journal of experimental child psychology*, 112(2), 208-230.
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, 4(8), 299-309.
- Severson, R. L., & Lemm, K. M. (2016). Kids See Human Too: Adapting an Individual Differences Measure of Anthropomorphism for a Child Sample. *Journal of Cognition and Development*, 17(1), 122-141. doi:<http://dx.doi.org/10.1080/15248372.2014.989445>
- Shimizu, Y., & Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental science*, 7(4), 425-430. doi:10.1111/j.1467-7687.2004.00362.x
- Stanton, A. (Writer). (2008). WALL·E [Motion Picture]. In. United States: Walt Disney Studios Motion Pictures.
- Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the national academy of sciences*, 106(43), 18362-18366. doi:<http://dx.doi.org/10.1073/pnas.0910063106>.
- Stekhoven, D. J. (2013). missForest: Nonparametric Missing Value Imputation using Random Forest. *R package version 1.4*.

- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York, NY: Basic Books.
- Victor, D. (2015). Hitchhiking Robot, Safe in Several Countries, Meets Its End in Philadelphia. *The New York Times*.
- Wei, C.-W., Hung, I.-C., Lee, L., & Chen, N.-S. (2011). A joyful classroom learning system with robot learning companion for children to learn mathematics multiplication. *TOJET: The Turkish Online Journal of Educational Technology*, 10(2), 11-23.
- Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the national academy of sciences*, 114(43), 11374-11379.
- Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford: Oxford University Press.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child development*, 72(3), 655-684. doi:<http://dx.doi.org/10.1111/1467-8624.00304>
- Wong, J. C. (2016). 'This is awful': robot can keep children occupied for hours without supervision. *The Guardian*. Retrieved from The Guardian website: <https://www.theguardian.com/technology/2016/sep/29/ipal-robot-childcare-robobusiness-san-jose>